

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:47:19 : Search time 1670.83 Seconds  
(without alignments)  
231.438 Million cell updates/sec

Title: US-09-396-196F-8  
Perfect score: 25  
Sequence: 1 gcatttcgacccctcgtcaggtgcag 25

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 773874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 182002

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_ba1:\*

2: gb\_ba2:\*

3: gb\_ba3:\*

4: gb\_in1:\*

5: gb\_in2:\*

6: gb\_in3:\*

7: gb\_om:\*

8: gb\_ov:\*

9: gb\_pat1:\*

10: gb\_pat2:\*

11: gb\_ph:\*

12: gb\_pl1:\*

13: gb\_pl2:\*

14: gb\_pl3:\*

15: gb\_pl4:\*

16: em\_ba1:\*

17: em\_ba2:\*

18: em\_fun:\*

19: em\_htgo\_hum:\*

20: em\_htgo\_inv:\*

21: em\_htgo\_rod:\*

22: em\_htgo\_hum1:\*

23: em\_htg\_hum2:\*

24: em\_htg\_hum3:\*

25: em\_htg\_hum4:\*

26: em\_htg\_hum5:\*

27: em\_htg\_hum6:\*

28: em\_htg\_hum7:\*

29: em\_htg\_hum8:\*

30: em\_htg\_inv1:\*

31: em\_htg\_inv2:\*

32: em\_htg\_other:\*

33: em\_htg\_rod:\*

34: em\_hum1:\*

35: em\_hum2:\*

36: em\_hum3:\*

37: em\_hum4:\*

38: em\_hum5:\*

39: em\_hum6:\*

40: em\_hum7:\*

41: em\_in:\*

42: em\_om:\*

43: em\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pl:\*

48: em\_ro:\*

49: em\_sts:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_vl:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_vl1:\*

59: gb\_vl2:\*

60: gb\_hcg1:\*

61: gb\_hcg2:\*

62: gb\_hcg3:\*

63: gb\_hcg4:\*

64: gb\_hcg5:\*

65: gb\_hcg6:\*

66: gb\_hcg7:\*

67: gb\_hcg8:\*

68: gb\_hcg9:\*

69: gb\_hcg10:\*

70: gb\_hcg11:\*

71: gb\_hcg12:\*

72: gb\_hcg13:\*

73: gb\_hcg14:\*

74: gb\_hcg15:\*

75: gb\_hcg16:\*

76: gb\_hcg17:\*

77: gb\_hcg18:\*

78: gb\_hcg19:\*

79: gb\_hcg20:\*

80: gb\_hcg21:\*

81: gb\_hcg22:\*

82: gb\_hcg23:\*

83: gb\_hcg24:\*

84: gb\_hcg25:\*

85: gb\_pr1:\*

86: gb\_pr2:\*

87: gb\_pr3:\*

88: gb\_pr4:\*

89: gb\_pr5:\*

90: gb\_pr6:\*

91: gb\_pr7:\*

92: gb\_pr8:\*

93: gb\_pr9:\*

94: gb\_r01:\*

95: gb\_r02:\*

96: gb\_in4:\*

97: gb\_pr10:\*

98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description               |
|------------|-------|-------------|--------|-------|---------------------------|
| 1          | 25    | 100.0       | 1041   | 9     | AR029499 Sequence         |
| 2          | 25    | 100.0       | 1041   | 9     | AR034916 Sequence         |
| 3          | 25    | 100.0       | 1084   | 9     | A11530 B10B gene o        |
| 4          | 25    | 100.0       | 1121   | 10    | E00893 Genomic DNA        |
| 5          | 25    | 100.0       | 5793   | 2     | ECOB10 J04423 E.coli 7,8- |
| 6          | 25    | 100.0       | 5872   | 9     | A38246 Sequence 1         |
| 7          | 25    | 100.0       | 5872   | 9     | A38251 Sequence 6         |
| 8          | 25    | 100.0       | 5872   | 9     | A93674 Sequence 1         |

| LOCUS                       | AR034916  | 1041 bp       | DNA                | PAT         | 29-SEP-1999  |
|-----------------------------|---|---------------|--------------------|-------------|--------------|
| DEFINITION                  | Sequence  | 7             | from patent        | US 5869719. |              |
| ACCESSION                   | AR034916  |               |                    |             |              |
| VERSION                     | AR034916.1  | GI:5950521    |                    |             |              |
| KEYWORDS                    | Unknown.  |               |                    |             |              |
| SOURCE                      | Unknown.  |               |                    |             |              |
| ORGANISM                    | Unclassified.                                     |               |                    |             |              |
| REFERENCE                   | 1 (bases 1 to 1041)                               |               |                    |             |              |
| AUTHORS                     | Patton,D.A.                                       |               |                    |             |              |
| TITLE                       | Transgenic plants having increased biotin content |               |                    |             |              |
| JOURNAL                     | Patent: US 5869719-A 7 09-FEB-1999;               |               |                    |             |              |
| FEATURES                    | Location/Qualifiers                               |               |                    |             |              |
| source                      | 1..1041   |               |                    |             |              |
|                             | /organism="unknown"                               |               |                    |             |              |
| BASE COUNT                  | 262 a   | 273 c         | 305 g              | 201 t       |              |
| ORIGIN                      |   |               |                    |             |              |
| Query Match                 |   | 100.0%;       | Score 25;          | DB 9;       | Length 1041; |
| Best Local Similarity       |   | 100.0%;       | Pred. NO. 2.6e-05; |             |              |
| Matches 25; Conservative 0; |   | Mismatches 0; | Indels 0;          | Gaps 0;     |              |
| OY 1                        | gcattcgatcctcgtcagtgacg 25                        |               |                    |             |              |
|                             |   |               |                    |             |              |
| Db 99                       | GCATTTCGATCCTCGTCAGTGCAG 123                      |               |                    |             |              |
| RESULT 3                    |   |               |                    |             |              |
| LOCUS                       | AR034916  | 1084 bp       | DNA                | PAT         | 09-FEB-1994  |
| DEFINITION                  | BioB gene of E.coli with primers.                 |               |                    |             |              |
| ACCESSION                   | AR034916  |               |                    |             |              |
| VERSION                     | AR034916.1  | GI:490218     |                    |             |              |
| KEYWORDS                    |   |               |                    |             |              |
| ORIGIN                      |   |               |                    |             |              |

ORGANISM

SOURCE

BASE COUNT  
ORIGIN

Query Mat  
Best Loca  
Match

19



```

/codon_start=1
/transl_table=11
/protein_id="AA023517.1"
/db_xref="GI:145427"
/translation="MATVKNQAIATAAFGRAAHYEQAHDQROSDALLMLPQRKYT
HVLDAAGCGPMWSRHRERHQAOTVALDLSPEMLVOAKQKDAADHYLAGDIESIPLATA
TFDLAMSLAVOWCGNLSLARELYRVAPRGVAFPTLVGSLPERHQAQVADDERP
HANRFLPDEIEOSLNGVYOHHTOPTLWDDALSMRSIKGIGATHLHGRDPRIL
TRSOLORLQALWPOQOGRPLTYHLFLGVIARE"
4938..5597
/gene="biob"
4938..5597
/gene="biob"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AA023518.1"
/db_xref="GI:145428"
/translation="MSKRYFVTGDTDEVGKTVASCALLQAKRAGRYTAGYKPVASG
EKTPGELRNSDALALQNSSLQIDYATVNTFAEPSPHITSAGEPRPESIVMSAG
LRALEQADWLVGVAGAGWETPLSDPTFPADWYQEQLPVILVYKVGCTINAMLLTA
OVIQHNAGLTLGAVANADVTPPKRHAFTMTTLRMIFRCWERSPGLQTKMRQPEES
T"
1554 a 1631 g 1245 t
BASE COUNT 1363 a 1554 c 1631 g 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccattcgcctcgtcagttcag 25
|||||
Db 2110 GCATTTCGATCCGTCGTCAGTGCGAG 2134

RESULT 6
A38246 A38246 5872 bp DNA PAT 05-MAR-1997
LOCUS Sequence 1 from Patent WO9408023.
DEFINITION A38246
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
Escherichia coli.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLES BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 1 14-APR-1994;
LONZA AG (CH)
COMMENT other publication PL 308301 950724
other publication CA 2145400 940414
other publication AU 4820293 940426
other publication HU 71781 960228
other publication SK 42095 951108
other publication CZ 9500809 950913
other publication FI 951547 950331
other publication JP 8501694T 960227.
FEATURES
source
1..5872
/location="Qualifiers"
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
/promoter
/function="PROMOTER PTAC"
/evidence="experimental"
23..28
-35_signal
45..49
-10_signal
/standard_name="PROMOTER PTAC"
/evidence="experimental"

RBS 105..119
/standard_name="BIOB RBS NO.9"
/evidence="experimental"
gene 117..1157
/evidence="experimental"
CDS 117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/number=1
/evidence="experimental"
/product="BIOTIN SYNTHASE"
/protein_id="CA002325.1"
/db_xref="GI:2294845"
/translation="MAHRPMTLSQVTELEKPLDLLEFAQOYVROHFPDROYVST
LISIKTAGPEDCKYCPOSSRYKGTLEAEERLMEVQVLESRAKKAAGSTPFGCAAN
KNRHERMDPYLEQWQVYKANGLEACMTLGTISEQARLNAGIDYTNHMLDTSPEP
VGNITLTTRTYORRLDTLEKVRAGIKVSGGIVGIGETVKORAGILLQALPTPES
VFINMLVKKGTPPLADNDVDVAFDEIRIIVARIIMPTSYVRLSGREONNEOTQAMC
FVAGANSLEYGCKLLTTPNPEDKDLQLEFRKILINPOOTAVIAGDNEQOQRLQALMT
PPTDEYVNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CA002326.1"
/db_xref="GI:2294846"
/translation="MATVKNQAIATAAFGRAAHYEQAHDQROSDALLMLPQRKYT
HVLDAAGCGPMWSRHRERHQAOTVALDLSPEMLVOAKQKDAADHYLAGDIESIPLATA
TFDLAMSLAVOWCGNLSLARELYRVAPRGVAFPTLVGSLPERHQAQVADDERP
HANRFLPDEIEOSLNGVYOHHTOPTLWDDALSMRSIKGIGATHLHGRDPRIL
TRSOLORLQALWPOQOGRPLTYHLFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence="experimental"
/product="DAPA SYNTHASE"
/protein_id="CA002327.1"
/db_xref="GI:2294847"
/translation="MTTDLAPDQRIHWPTYSPLPYVYVSAEGCELLISDGR
LVQMSMMAAIGHVNPOLNAAKSOIDAMSHVMGCTHAPALCIRIVAMPPOP
LECVFLADSGVAEVMKMALOYQAKGARORFLTPNGYGDGREGMSVCDPNS
MHSIMKGYLPENLEAPAPOSRMDGEMDERMVFALMAHREIAAVITEPLVQAG
GMKRYHPMLKRIKICDRGILLIDELATGFGRTGLFACCHAEIADPILICLQAL
TGGTMTLSATLTTREVAETISNGEACFMHGPFPKPNPLCAANAASLAILSGDQQ
QVADIEVQLEQALPARDAEMVADVVLGALGVETTHPVNMAALQKFVEQGVWIRP
EGRLIILMPYIILLPOOLRLFAVAVRAVODEFFEQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```



/evidence="experimental"  
/product="PROTEIN"  
/protein\_id="CAA02328.1"  
/db\_xref="GI:2294848"  
/translation="MKLISNDLRDGLKPHRHVFNMGVGDGNISPHLAMDDVPAGTK  
SVVTCYDDADAPGSGMHWVYNLPADRVLPDGGSGSLVAMPDGVLOTFRDPKGTG  
YDGAAPPKGETHRYIFTVALDIERIDVEGASGAWVGFVNFHSLASITAMFS"  
stem\_loop  
5583..5605  
5583..5644  
/standard\_name="RHO-INDEPENDENT TRANSCRIPTIONAL  
TERMINATOR"  
BASE COUNT 1318 a 1552 c 1695 g 1307 t  
ORIGIN  
Query Match 100.0%; Score 25; DB 9; Length 5872;  
Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gcattcgatccctgcaggtgcag 25  
|||||  
Db 215 GCATTGATCCTCGTCAGGTGCAG 239  
RESULT 7  
A38251 5872 bp DNA PAT 05-MAR-1997  
LOCUS  
DEFINITION Sequence 6 from Patent WO9408023.  
A38251  
ACCESSION  
VERSION A38251.1 GI:2294849  
KEYWORDS  
SOURCE  
ORGANISM  
Escherichia coli.  
Escherichia coli  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.  
REFERENCE 1 (bases 1 to 5872)  
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.  
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN  
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;  
COMMENT  
LONZA AG (CH)  
Other publication PL 308301 950724  
Other publication CA 2145400 940414  
Other publication AU 4820293 940426  
Other publication HU 71781 960228  
Other publication SK 42095 951108  
Other publication CZ 950809 950913  
Other publication FI 951547 950331  
Other publication JP 8501694T 960227.  
FEATURES  
source  
1. 5872  
Location/Qualifiers  
/organism="Escherichia coli"  
/strain="DSM498"  
/db\_xref="taxon:562"  
/clone="PBO30A15-9"  
1141..1156  
/standard\_name="BIOF RBS"  
1154..2308  
/gene="BIOF"  
1154..2308  
/gene="BIOF"  
/gene="BIOF"  
/standard\_name="8-AMINO-7-OXONONANOATE SYNTHASE"  
/EC\_number="2.3.1.47"  
/codon\_start=1  
/transl\_table=11  
/number=2  
/evidence="experimental"  
/product="KAPA SYNTHASE"  
/protein\_id="CAA02329.1"  
/db\_xref="GI:2294850"  
/translation="MSMOEKINAAIDARADALRRYPVAGAGRMIVADRDYVNF  
SNDYLGSHHPOTIRAMOGAEQDPTGSGSGHSGSVYHOALEEELAMWLGYSRA  
LIFISGFAANQAVIAAMAKEDRIADRLSHASLEASLSFSLKRRANDVTHLAR  
ILASPCGQMVYVEGVFSMDGDSAPLAEIOQVTOOHNGWMLVDDAHGTGVIAGEGRC

SCMIQKVELLYVTFGKFGVSGAAVLCSTVADYLLQFARHLIYSTSMPPAQAL  
RASLAVIRSDGDAAREKIALITIRFAGVODLPFTLADSCSALOPLIVGDNRAIOL  
AEKEBOQGCWTAIRPPVPAGTARLRILITLAHEMODIDRLLEVLHNGC"  
RBS  
3030..3045  
/standard\_name="BIOB RBS"  
3043..3753  
/gene="BIOB"  
3043..3753  
/gene="BIOB"  
/gene="BIOB"  
/standard\_name="DETHIOBIOTIN SYNTHASE"  
/EC\_number="6.3.3.3"  
/codon\_start=1  
/transl\_table=11  
/number=4  
/evidence="experimental"  
/product="DTB SYNTHASE"  
/protein\_id="CAA02330.1"  
/db\_xref="GI:2294851"  
/translation="MSKRYFVYGDTEVGTAVASCALLQAKAAGYRTAGYKPVASGS  
EKTPEGLRSDALORNSSLQIDYATVNPYFAEPSPHITSAQEGRPISLVMSAG  
LRALBOADWVIVEGAGGMPPLSDTPFADWVVOEOLPVLVGVKLGJNHMLTA  
OYTORAGLTIAGVANDYTPPKRIAEYNTLTIRIPARPLGELPMLAENENATGK  
YINLAFVDASTLGTSTRL"  
BASE COUNT 1318 a 1552 c 1695 g 1307 t  
ORIGIN  
Query Match 100.0%; Score 25; DB 9; Length 5872;  
Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gcattcgatccctgcaggtgcag 25  
|||||  
Db 215 GCATTGATCCTCGTCAGGTGCAG 239  
RESULT 8  
A93674 5872 bp DNA PAT 22-JAN-2000  
LOCUS  
DEFINITION Sequence 1 from Patent EP0798384.  
A93674  
ACCESSION  
VERSION A93674.1 GI:6741862  
KEYWORDS  
SOURCE  
ORGANISM  
Escherichia coli.  
Escherichia coli  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.  
REFERENCE 1 (bases 1 to 5872)  
AUTHORS Birch, O. and Brass, J.  
TITLE Biotechnological method of producing biotin  
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;  
COMMENT  
LONZA AG (CH)  
FEATURES  
source  
1. 5872  
Location/Qualifiers  
/organism="Escherichia coli"  
/strain="DSM498"  
/db\_xref="taxon:562"  
/clone="PBO30A-15/9"  
1..96  
/function="PROMOTER PTAC"  
/evidence="experimental"  
23..28  
/standard\_name="PROMOTER PTAC"  
/evidence="experimental"  
105..119  
/standard\_name="BIOB RBS NO.9"  
/evidence="experimental"  
117..1157  
/gene="BIOB"  
117..1157  
/gene="BIOB"  
/gene="BIOB"  
/codon\_start=1

```

/transl_table=11
/evidence=experimental
/product_id="BIOTIN SYNTHASE"
/db_xref="GI:6741863"
/translation="MAHRPWTLSQVTELEFELKPLDLLFEAOOVHROHPDPOVOST
LLSKTGPCECKCYCPOSSRYKGLAEERLMEVEQVLESRAKKAAGSTRFCGAAV
KNPHERMPYELQVOCYKKAAGLACMTLGLTSSQAKRIANGGLDYNNLPTSPEF
VGNITTRTVOERLDTLEKVDAGIKVCSGIVGLGETVDRGALLQIANLPTPES
VPIMLVKVGTPLADNDVDADPEIRFTIYAVRIMPTSYVRLSAGREOMQTOAMC
FMAGANSIFYGCKLTTPNPEEDKDLQIFRKILNPQOTAVLAGDNPOQRLPQALIT
PDIDEYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKNQALAAFGRAAHYEQHADLROSADALLAMPORKYT
HVLDAAGCGPMGMSRMRERHAYOYTAIDLSPPMTYOROKDADHYLAGTIESLPLTA
TFLDAMNLAVMCSGNLSTALRELYRVVRKGVATTTVOGSLPELHOAWQVDRP
HANRFLPPDEILOSINCVHIOHTIPTLMEFDALSAMSLKIGATHLHEGRDPL
TRSOQLRQLQAWPOQGRYPLTYHLFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDDLAFDORHIMHPYTSMTPLVPVVSABGCELLSPGR
LVDMSSMAALHGVNHPOLNAAMKSIDAMSHWPGGTTTAPATELCKVAMTPOR
LECVLADSGSAVAYAMKMAIQYMAKGEARORPLTFRNGIHGDTFGMSVCDPDS
MSLAKGCTPENNIPAPPOSRMDEGDEWGFARLMAHREHIAAVYIEPILVQAG
GMATHEPWLKRIKICDREGILLADELATGFGTKLACEHAIEADILGLKAL
TGGTMTLSATLTTRVAAETISNGEAGCEPMHGPTFMGNPLACANAASALIESGMOO
OVADIEVLRQDLAPARDAEMVADVGAIGVETTHPVNMAALQKFFVEGCVIIRP
PKLIYMPPTIILPQOLRLTAAVNRAVQDETFFCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDQKLPHRHVFNGMGYDGDNISPLHAMDVAGTK
SEVVYCDPADPAGSGMWVYVNPATRYLPOGFGSLVAMDGVLIQTDTDRKTS
YDGAAPKGETHRTFYVHALDIEKIDVBSAGAMGVNHFHSLASASTAMFS"
5583..5605
stem_loop
terminator
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1552..1695
BASE COUNT 1318 a 1552 c 1695 g 1307 t

```

```

ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 gcatlccatccctcagtcagtcag 25
Db 215 GCATTTCGATCCCTCGTCAGSTGCAG 239

RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O. and Brass,U.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
location=Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/cclone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQKRIINALDARRAADALRRRYPAQAGKMLVADPROLYNF
SSNXYCLSHHPQITIRAWOQAGDFGSGSGSHVSGTSYVHQALESLEAELEWIGYSRA
LFTISGFANQAVYAAAMAKEDRTIADRLSHASLEASLSPQLRFAHNDVTHLAR
LLASPCGQGVVTEGVFSMDGDSAPLAETIQVTOOHNGSLAMDADHAGTGIVGEGRG
SCWLQKVPBLVVTTEGKFGVSGAVALCSTVADLLQFARHLIYSTGMPAQAOL
RASLAVIRDEGDARREKLAALITFRPAGVQDLPFLADSCSAIOPLIQDNRALQ
AEKLRQGCWVTIARPTVPAGTARLRLITTAHHEQDDIRLEVLHGNG"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DDB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFVGTDTDEVKTYVASCALLQAAAGTRTATGYTPVAVSGS
EKTPESLRNSDALQDRSSLDQDTATVNPYTFPAEPTSPHILISADGRFIESLVNSAG
LRALQOADMVNLVEGAGWFTPLSDTFEADWVTOEQLVILVGVCKLGCINHAMLTA
QVIOHAGLTLAGVANDVTPPGKRHAIEYVTTLTRMIPAPLLGLCEIPLAENPENATYK

```

YINLAFVDASTLIGFTSRL"  
 BASE COUNT 1318 a 1552 c 1695 g 1307 t  
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatctgcattcctgcagtgacg 25  
 Db 215 GCATTTCGATCCTCGTCAGTGACG 239

RESULT 10  
 LOCUS AR101809 5872 bp DNA PAT 14-FEB-2001  
 DEFINITION Sequence 1 from patent US 6083712.  
 ACCESSION AR101809  
 VERSION AR101809.1 GI:12812607  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)  
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.  
 TITLE Biotechnological method of producing biotin  
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;  
 FEATURES Location/Qualifiers  
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t  
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatctgcattcctgcagtgacg 25  
 Db 215 GCATTTCGATCCTCGTCAGTGACG 239

RESULT 11  
 LOCUS AR101810 5872 bp DNA PAT 14-FEB-2001  
 DEFINITION Sequence 6 from patent US 6083712.  
 ACCESSION AR101810  
 VERSION AR101810.1 GI:12812608  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)  
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.  
 TITLE Biotechnological method of producing biotin  
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;  
 FEATURES Location/Qualifiers  
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t  
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatctgcattcctgcagtgacg 25  
 Db 215 GCATTTCGATCCTCGTCAGTGACG 239

RESULT 12  
 LOCUS AE000180  
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.

ACCESSION AE000180  
 VERSION AE000180.1 GI:1786988  
 KEYWORDS  
 SOURCE Escherichia coli K12.  
 ORGANISM Escherichia coli K12.  
 Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 11022)  
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.  
 TITLE The complete genome sequence of Escherichia coli K-12  
 JOURNAL Science 277 (5331), 1453-1474 (1997)  
 MEDLINE 97426617  
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)  
 AUTHORS Blattner, F.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 3 (bases 1 to 11022)  
 AUTHORS Blattner, F.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 4 (bases 1 to 11022)  
 AUTHORS Plunkett, G. III.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 This sequence was determined by the E. coli Genome project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@amherst.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). \*\*\* The E. coli K12 sequence and its annotations are periodically updated: this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'p' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES  
 source

1..11022  
 /organism="Escherichia coli K12"  
 /strain="K12"  
 /sub\_strain="MG1655"



```

VERSION      AE005258.1 GI:12513751
KEYWORDS     .
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
              Nature 409 (6819), 529-533 (2001)
TITLE        Nature 409 (6819), 529-533 (2001)
JOURNAL      21074935
MEDLINE      11206551
PUBMED
REFERENCE
AUTHORS
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-Oct-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. 13501
   /organism="Escherichia coli O157:H7 EDL933"
   /strain="EDL933"
   /serotype="O157:H7"
   /db_xref="taxon:155864"
   /note="enterohaemorrhagic"
   <1..7576
   /note="O-island #36: Region of the EDL933 chromosome not
   homologous to E. coli K-12 MG1655: Cryptic prophage
   CP-933K; Includes one copy of the 13 bp direct repeat that
   flanks the prophage"
   66..665
   /gene="10mK"
   /note="20981"
   66..665
   /gene="10mK"
   /function="putative membrane; other or unknown (Phage or
   prophage Related)"
   /note="Residues 1 to 199 of 199 are 72.86 pct identical to
   residues 1 to 199 of 199 from Genpept 118 :
   g1175327891gb|AA63231.1|AF151091_2 (AF151091) Lom
   [Prophage P-Elba]"
   /codon_start=1
   /transl_table=1
   /product="putative outer membrane protein of prophage
   CP-933K"
   /protein_id="AA65137.1"
   /db_xref="GI:12513752"
   /translation="MRKVCALLSAICLAVSGVPAMASRHOSTLSAGYIHASTDAPG
   SDDLINVKRYREFTDTLGLITSFSAANEDEOKHYSDTRWHEDEVNRKRWSVAG
   PSYVNMESFAYAMAGVAYSRSVTFSGDYFRVNDNKRKTHDVLTGSDARYSNTSLAM
   GAGVQNPFTESAVDAVAYEXSGDMRTGDFIVGVYKF"
   719..2041
   /gene="20982"
   /note="20982"
   719..2041
   /function="putative structure; structural component (Phage
   or prophage Related)"
   /note="Residues 164 to 440 of 440 are 68.79 pct identical
   to residues 381 to 645 of 645 from Genpept 118 :
   g145854361gb|AAD25464.1|AF125520_59 (AF125520) putative
   tail fiber protein [Bacteriophage 933W]"
   /codon_start=1
   /transl_table=1
   /product="putative tail component of prophage CP-933K"

   /protein_id="AA65138.1"
   /db_xref="GI:12513753"
   /translation="MGMAVQISGVCKDAGKPTQCTQIOLAKRRNSTVYVNTASE
   NPDEAGRYSMDEVETGOYSTVLVEGPPPSHAGTISYEDBSQETLNDPLIGAMREDDAR
   PEALRAREQVVEEAARHAEEAKKNAGAEFTSARNAGISAKASASANADTSADSE
   SARQARREASAAAKSESESSSSAEEAOKSESLQSTADSELSKRTAEASNAADA
   TTSERKARESAESKQSAEQRIAEDAVNRIPVYVPGPGKGEPPGQPGKGDGE
   RGDPGATGATGERGGDGTGPAGPGKPGRGEGTGLGNAGPGPGKGDGAGPA
   GPQPKKETGACAGVAGTACPGKPGKPGDGEFTQIFRIGPMRIITNSYNGPCDGLI
   TGLTFDLPKDATVOVGFMHQLVRFQFGPGQMDVKGLGDEVSDFGRIGE"
   2004..2312
   /gene="20984"
   /note="20984"
   2004..2312
   /gene="20984"
   /function="orf; Other or unknown (Phage or Prophage
   Related)"
   /note="Residues 1 to 102 of 102 are 98.03 pct identical to
   residues 28 to 129 of 129 from Genpept 118 :
   g145854371gb|AAD25465.1|AF125520_60 (AF125520)
   hypothetical protein [Bacteriophage 933W]"
   /codon_start=1
   /transl_table=1
   /product="unknown protein encoded by prophage CP-933K"
   /protein_id="AA65139.1"
   /db_xref="GI:12513754"
   /translation="MKMAVIOALELNDMNLIRKIMQSLCGGKHDDCENGSRSLTAOLR
   LGPADLLESDNGIIPEDRVITGVVLADAKKQIQCVVRPLQILRADGTWENIGMK"

   2489..3469
   /gene="20985"
   2489..3469
   /gene="20985"
   /function="orf; Other or unknown (Phage or Prophage
   Related)"
   /note="Residues 13 to 326 of 326 are 55.73 pct identical
   to residues 18 to 331 of 336 from Genpept 118 :
   g1169603671gb|AA633527.11 (AF170176) hypothetical protein
   predicted by Glimmer [Salmonella typhimurium L72]"
   /codon_start=1
   /transl_table=1
   /product="unknown protein encoded by prophage CP-933K"
   /protein_id="AA65140.1"
   /db_xref="GI:12513755"
   /translation="MLSPIRTFHNSVNIYVSSPCQVTSFAGKEVELKVIDEPTILE
   OMFEENPERKDEVPYIVNTKOPHLDVNTAARLESMDMIGIEVDGDFSVNOKTAS
   KLERPDENVMIITREDVDSPMDRKLSDIYHDIIOEORLRTEDKDEYLLMLKEIR
   ETSKQDSLISMTAKRRNHAMDFERNALILKAGEIFPCTYNTKPHGTSFGEGCYLD
   MMILLTGKLGITVADPGISMHYDRNDSVNIENSAIIVNRSRHPALLEGLSPMISKYD
   AHPYVDLGLGKGVKYYFNFTPLNHNHDFIEFNHPIIMNTSQTCSW"
   3530..4522
   /gene="20986"
   3530..4522
   /gene="20986"
   /function="orf; Other or unknown (Phage or Prophage
   Related)"
   /note="No significant matches"
   /transl_table=1
   /product="unknown protein encoded by prophage CP-933K"
   /protein_id="AA65141.1"
   /db_xref="GI:12513756"
   /translation="MKIPSLQSNFNSAPAGSADIPAPRAENAYADVYLIGKRIPL
   /SAADLSNVESEVYRAVHDSRSLIDQHTVDMIGNVLDLSQFFRAVSAGINEK
   VHTGCIKYRNEELNNESSVKIIDDQSLCNELLYSDGQOEIPFICAGENDNEEPY
   VSRVAPPDSDYEMSGKGIHEIILHHTGSSDSGSDNIEIGTEILARVAOELG
   WSPVDPKGYAEPREHNLRLNLNALRQAMHENEERAFPERLGTISRYEASPDFT
   EYSVASNIGYGFQOHDFGLAINDLQDANOIQLHCAPIYFTGVDKHNQO"
   5350..6231
   /gene="20989"
   5350..6231
   /gene="20989"
   /function="orf; Other or unknown (Phage or Prophage
   Related)"

```

```

/note="No significant matches"
/codon_start=1
/translation="unknown protein encoded by prophage CP-933K"
/product="GI:12513757"
/db_xref="GI:12513757"
/translation="MSPYSVNIAGCSNSTRNLSPDNKVLSSVDAVHSDNGAQQ
KAGNRYVATDNKFCVRESHSGCTNLHRLGMPKGEISRIEVMANASVSAAM
ERGIHSHNRPDLPPVAPPELPSVDYVNRISVGNVIGKGNVAVVEDATKVLK
FTTOSNEEVTSVRCFNQYAGASAEK IYGNNGDILIGIMDK INESDLINISLP
AEHAIYDMFRLERKGIILFVDTETNVLDRANKENPIDISSVNSDSMSQIMO
SYHGKODLSVLSKTI"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
/function="orf; other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/translation="unknown protein encoded by prophage CP-933K"
/product="GI:12513758"
/db_xref="GI:12513758"
/translation="MRPTSLNLVLIHOSRSSMSDPDIRSLKAVSQMIRKNPOLR
OGDHNIIYOIEALDPIKISTENGRLVLAISIRISRIKSEYVVIHINSRIGVMAHRD
IDAEHNRGSGDFHCLNAVEYPCGSGISVVPFHATVYHLLHVFHNLNGERIKVES
SRAESOKYSPILLEPARIYGLAFSEEVLSENKFHEIGMPRTSPYDLSLHDDNT
VSLGFOOVRHLPL"
complement(7668..8144)
/gene="ydhb"
/note="20992"
complement(7668..8144)
/gene="ydhb"
/function="orf; unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/translation="orf; hypothetical protein"
/product="orf; hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGDLPKHHVFMGMDYDGNISPLHAMPDVPAQFK
SFVAVCYDPDAPRTGSGMHWVYVNLPAQFSGSLVAMPDGVLOTGPDGKTC
DGAAPPKGETRIYFTYHALDVERIDYDEGASGAMVGFNVHPSLASASTTAMFS"
complement(8203..9492)
/gene="dica"
/note="20993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatctgactctctcagtcagcag 25
|||||
Db 9677 GCATTTCGATCCTCGTCAGTGTGAG 9701

RESULT 14
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)
/note="No significant matches"
/codon_start=1
/translation="unknown protein encoded by prophage CP-933K"
/product="GI:12513757"
/db_xref="GI:12513757"
/translation="MSPYSVNIAGCSNSTRNLSPDNKVLSSVDAVHSDNGAQQ
KAGNRYVATDNKFCVRESHSGCTNLHRLGMPKGEISRIEVMANASVSAAM
ERGIHSHNRPDLPPVAPPELPSVDYVNRISVGNVIGKGNVAVVEDATKVLK
FTTOSNEEVTSVRCFNQYAGASAEK IYGNNGDILIGIMDK INESDLINISLP
AEHAIYDMFRLERKGIILFVDTETNVLDRANKENPIDISSVNSDSMSQIMO
SYHGKODLSVLSKTI"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
/function="orf; other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/translation="unknown protein encoded by prophage CP-933K"
/product="GI:12513758"
/db_xref="GI:12513758"
/translation="MRPTSLNLVLIHOSRSSMSDPDIRSLKAVSQMIRKNPOLR
OGDHNIIYOIEALDPIKISTENGRLVLAISIRISRIKSEYVVIHINSRIGVMAHRD
IDAEHNRGSGDFHCLNAVEYPCGSGISVVPFHATVYHLLHVFHNLNGERIKVES
SRAESOKYSPILLEPARIYGLAFSEEVLSENKFHEIGMPRTSPYDLSLHDDNT
VSLGFOOVRHLPL"
complement(7668..8144)
/gene="ydhb"
/note="20992"
complement(7668..8144)
/gene="ydhb"
/function="orf; unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/translation="orf; hypothetical protein"
/product="orf; hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGDLPKHHVFMGMDYDGNISPLHAMPDVPAQFK
SFVAVCYDPDAPRTGSGMHWVYVNLPAQFSGSLVAMPDGVLOTGPDGKTC
DGAAPPKGETRIYFTYHALDVERIDYDEGASGAMVGFNVHPSLASASTTAMFS"
complement(8203..9492)
/gene="dica"
/note="20993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatctgactctctcagtcagcag 25
|||||
Db 9677 GCATTTCGATCCTCGTCAGTGTGAG 9701

AUTHORS
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C., and Shinagawa, H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
2 (sites)
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)
Yokoyama, K., Makino, K., Kubota, Y., Matanabe, M., Kimura, S.,
Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsuka, E., Nakayama, K., Murata, T.,
Tanaka, M., Ito, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 297816)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: kenegen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
genome project.
Location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECS0753"
79..1245
/gene="ECS0753"
/note="ECS0753"
/note="similar to SUCC_ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/translation="orf; hypothetical protein"
/product="orf; hypothetical protein"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="MNLHEVQAKQFARYGLPAPGYACTPPREAEASKIGAGPW
VRCQVHAGRGKAGVYVNSKEDIRAFKEMNLGRLVYQDNDGQPVNQTIVPAAT
DIKELIYGAAYVDRSSRYVPMASREGVELEKVAERPHILHKAALDPLTPMPYQG
RELAKRIGLEGIVQOFTKIPMGALTIFLERDLALITENPLVITKQGLICLDKELGA
DGNALRPPDLREMRDQSOEDPREAQAWQELMVALDGNIGCVNAGGLMGMDIV
KLHGEPAFNLFDVGGGATKERVTEAFKIIISDQKAVLVNIFGIVGCDLIAGGIIG
AAAEVGVNVPVVRLEGNNALGAKKLADSGILITIAKGLTDAAGQVVAAVEAG"

```

```

gene
1245..2114
/gene="ECS0754"
CDS
1245..2114
/gene="ECS0754"
/note="similar to SUCD_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/gene="ECS0755"
2781..3686
/gene="ECS0755"
2781..3686
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator)"
Escherichia coli g1417043|sp|P32064|GQVA_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/transl_table=11
/note="probable cob(I)alamin adenosyltransferase, similar
to cob(I)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
g11151481|sp|P13040|BtUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/transl_table=11
/note="MEARISTERHOOROKLEPOVTRVAALTEKKGILVETGNGK
KSTRAAGTAVRAVGHGKTVGAYKIGMDNGEYVNLQIAGVPHMGSGFWETONR
QADIDAKREVWSSKRLADKRDIVYDEITVMLAYHLDDEVIASLQNRPAQGV
IVTGRGCHSILKMDVSEIRPVKHAFDGIDQAPGIDW"
complement(4332..5984)
/gene="ECS0757"
4332..5984
/complement(4332..5984)
/gene="ECS0757"
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
g11205981|sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

```

```

/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/transl_table=11
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g1121467|sp|P24943|GUTR_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34182.1"
/gene="ECS0759"
7533..7850
/complement(7533..7850)
/gene="ECS0759"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

Query Match 100.0%; Score 25; DB 2; Length 297816;  
Best Local Similarity 100.0%; Pred. No. 2,3e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaattcgatccctcgtaagtgacag 25  
|||||  
db 92409 GCATTCGATCCTCGTCAGGTCAG 92433  
|||||

RESULT 15  
ECOBIOB 128 bp DNA BCT 03-JAN-1995  
Escherichia coli biotin (biob) gene, early terminator region.  
M27731  
M27731.1 GI:341755  
biob gene: biotin.  
Escherichia coli (strain K-12) DNA.  
Escherichia coli  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.  
1 (bases 1 to 128)  
Nath, S. K.  
Attenuation of transcription of biotin genes in Escherichia coli  
Can. J. Microbiol. 34 (12), 1288-1296 (1988)  
89167942  
Location/Qualifiers  
1..128  
/organism="Escherichia coli"  
/strain="K-12"

```

gene
  /db_xref="taxon:562"
  4..128
  /gene="b10B"
-10_signal
  4..9
  /gene="b10B"
  20..>128
  /gene="b10B"
  103..121
  /gene="b10B"
  /note="early terminator"
BASE COUNT      31 a      28 c      34 g      35 t
ORIGIN

Query Match
Best Local Similarity 84.0%; Score 21; DB 2; Length 128;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gcattcgatcctcgcaggt 21
        |||
Db       59 GCATTTCGATCTCTGTCAGGT 79

RESULT 16
AF250776      5526 bp      DNA      BCT      31-JAN-2001
LOCUS
DEFINITION
  uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
  modc-b10A intergenic region, DAPA-aminotransferase BioA (b10A),
  biotin synthase BioB (b10B), KAPA synthetase BioF (b10F), and
  dehydrobiotin synthetase BioD (b10D) genes, complete cds; and
  AF250776
  AF250776.1 GI:12620124

ACCESSION
VERSION
KEYWORDS
SOURCE
  uncultured bacterium pCosHE2.
  Bacteria; environmental samples.
ORGANISM
REFERENCE
  1 (bases 1 to 5526)
  Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streif,W.R.
  Direct cloning from enrichment cultures: a reliable strategy for
  isolation of complete operons and genes from microbial consortia
  Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
JOURNAL
  20575196
  11133432
  2 (bases 1 to 5526)
  Entcheva,P., Liebl,W. and Streif,W.R.
  Direct Submission
  Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
  Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
  source
  1..5526
  /organism="uncultured bacterium pCosHE2"
  /db_xref="taxon:143797"
  /clone="pCosHE2"
  /note="unknown organism, cosmid clone derived from
  environmental consortium"
  complement(52..528)
  /note="ORF1"
  /codon_start=1
  /transl_table=11
  /product="hypothetical 17.1 kDa protein in modc-b10A
  intergenic region"
  /protein_id="AAG60577.1"
  /db_xref="GI:12620125"
  /translation="MKLISNDLRDGLKLPRIHVFNGMGYDGNISPHLAMVDYPAQTK
  SFVVTGYDPAIPGSGWMMHVVVNLPAIDRVLPDGGSLVAMPQCVLOTRTDFKAG
  YDGAAPKPGKHRIPTVHALDIERLDIVDEGASGAVNGVNFHFSLASASTAMPS"
  complement(587..1876)
  /gene="b10A"
  complement(587..1876)
  /gene="b10A"
  /note="7,8-diaminopelargonic acid
  synthetase-aminotransferase"

gene
  /codon_start=1
  /transl_table=11
  /product="DAPA-aminotransferase BioA"
  /protein_id="AAG60578.1"
  /db_xref="GI:12620126"
  /translation="MTTDDLPADPDRHIMHPYNSMSPLPYVYVASECEIILSDGR
  LVDSMSMAAIIHGYNHPOLNMAKSDIAMSRYVFGCITAPALIEICRLKLVAMPPOP
  LKCVPLADSGSAVEYAMKALQYQAKCEARLITRNIGHDTFGAMSVCDPNS
  MSLKMGYLPENLIFAPQSRMDGEMDERDYGAFARLMAARHHEIAVIEPIYQAG
  GMRWYPEWMLKRIKICDREGILLADIELATGFGRTGKLFACHEIAEIAIDILGKAL
  TGGIMTLISATLTTRREVAETISNGEAGCEFWHGTFMGNDPLACAANAASIALIESGDMO
  QVADIEVOLREOLAPADAEMADVAVGALGIVETTPVMAALQKRFVQGWITP
  FGLIYLPMPYITILPQOLRLTAAVNRAVQDETFECQ"
  1963..3003
  /gene="b10B"
  1963..3003
  /gene="b10B"
  /codon_start=1
  /transl_table=11
  /product="biotin synthase BioB"
  /protein_id="AAG60579.1"
  /db_xref="GI:12620127"
  /translation="MAHRRMTLSQVTELEFKPLLDLFEAQVYHQHPDPROYST
  LLSIKTAGCEDECKYCPQSSRYKGLERLEMEYQVLESARKAKAGSRFCGGAAM
  KNPNERMDPELYEOMVQKDLGLEACMTLGLSSQORLANGLDYNNILDPSPER
  YNNITTTTYOERLDTLEKVRDAGIKVCSGGIVGLGFTVKDRAGILLQIANLPPEPS
  VPIMLVYKGTPIADNDVDPAFPIRTIAVARIIMPYSYRISAGCEQMEQTOAC
  FMAAGNSIFGCGKLLTTPNPEDDLQIFRKIGINPQOTAVLADGNEQORLBOALMT
  PPTDEYNAAL"
  3000..4134
  /gene="b10F"
  3000..4134
  /gene="b10F"
  /note="8-amino-7-oxononanoate synthase"
  /codon_start=1
  /transl_table=11
  /product="KAPA synthetase BioF"
  /protein_id="AAG60580.1"
  /db_xref="GI:12620128"
  /translation="MSMOEKINALDARRAADLRRLRRYPVAGAGRIIVADPROYINP
  SSNDYLGSHHPOLIRAMOCGAEQFGISGSGVSGSYVYHQBALBEELEBWLIGYSRA
  LLFTSGFANQAVATAAMAKEDRTAARLSHASLSEASLSPQSLRRFHAHNDYTHLAR
  LLAPRCBOOLVYTEGYFSMDGSAPIAEIQQVYQNGNLAWDADHAGVIGEORG
  SCWLOKAPBELLVTEFGKFGVSGAAVLCSTVADYLDFARHLISTGMPPAQOAL
  RASLAVIRDEGDARRKELVSLAHFRAGVQDLPFTLADSCSAIOPLIYDMSBALQL
  AEKLROOQCVWTARIPPTVPAFTARLRLTLTAHEMQDIDRLLEVLGNG"
  4141..4896
  /gene="b10C"
  4141..4896
  /gene="b10C"
  /note="reaction step prior to synthesis of pimeloyl-CoA"
  /codon_start=1
  /transl_table=11
  /product="biotin biosynthesis protein BioC"
  /protein_id="AAG60581.1"
  /db_xref="GI:12620129"
  /translation="MATVKNQATAAAFGAALAYEQHADLRQSDADVLLAMPERRKYT
  HVDAGCGPGWMTRMHMERHAQVYALDLSPPMIVQARQDADHYLAGIDIESLPLATA
  TFDLAMSNIAMWOCNLSALRELYRVVRSKGVAFETTLVQSLPELHOAMQAVERP
  HANFLPEPDEIEOSLNGVYOHIIQPTITLMPDASAMNSLKGIGATIHHEBRDRIL
  TRSOLRLQALAMPQOQGRPLVYHFLGVIANE"
  4883..>5526
  /gene="b10D"
  4883..>5526
  /gene="b10D"
  /note="DTB synthetase"
  /codon_start=1
  /transl_table=11
  /product="dehydrobiotin synthetase BioD"
  /protein_id="AAG60582.1"
  /db_xref="GI:12620130"
  /translation="MLVSKRYFVTGTDREVGKTVASCALLOAKAAGRTAGYKPVAS
  GSEKTPEGIRNSDALALGRNSSSLQIDYATVNPYTFAPETPSHIISAGGRPIESSVNS

```



SCRLAEQADWVIEGAGFPTPSDTFTPADWTOEOLPVLLVWGKICGINHML  
TAQAIQHAGLTLAGWANDVPPKGRHAETFTTLRIMPALGEITWLANEP"

BASE COUNT 1274 a 1507 c 1567 g 1178 t

ORIGIN

Query Match 84.0%; Score 21; DB 2; Length 5526;  
Best Local Similarity 100.0%; Pred. NO. 0.0083;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatttcgacccgcgcaggt 21  
|||||

Db 2061 GCATTTGCATCTGCTCAGGT 2081

# RESULT 17

LOCUS MTC1429B 2980 bp DNA BCT 17-JUN-1998  
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 38/162.  
ACCESSION Z96797 AL123436  
VERSION 296797.1 GI:3261811  
KEYWORDS  
SOURCE Mycobacterium tuberculosis.  
ORGANISM Mycobacterium tuberculosis.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;  
Mycobacterium; Mycobacterium tuberculosis complex.  
1 (bases 1 to 2980)  
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,  
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,  
Tejeda, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
Comor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,  
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
Squires, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and  
Barrell, B.G.

# REFERENCE

Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence  
Nature 393 (6685), 537-544 (1998)  
Erratum: [[published erratum appears in Nature 1998 Nov  
12;396(6707):190]]  
2 (bases 1 to 2980)  
Parkhill, J.  
Direct Submission  
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium  
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique  
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
75124 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
On Jun 27, 1998 this sequence version replaced gi:2190471.

# REFERENCE

Notes:  
Details of M. tuberculosis sequencing at the Sanger Centre are  
available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)) CDS have  
been renumbered from the original cosmid submissions but the old  
gene designations are in brackets after the new gene numbers.  
Gene prediction was based on a Hidden Markov Model of TB genes  
implemented in TParse (Krogh) supplemented with visual inspection  
of positional base preference in codons, especially where there is  
an increase in the observed/expected third position G + C.  
CAUTION: In some cases we may not have predicted the correct  
initiation codon. Where possible we choose an initiation codon  
(atg, gtg, or ttg) which is preceded by an upstream ribosome  
binding site sequence (optimally 5-13bp before the initiation  
codon). If this cannot be identified we choose the most upstream  
initiation codon.

# COMMENT

# FEATURES

Source  
1..2980  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"

misc\_feature

misc\_feature

misc\_feature

repeat\_unit

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

<1..613  
/note="fragment designated v042. Does not represent a  
physical clone"

/complement(11..38)  
/note="inverted repeat at 3'-end of IS6110"

39..42  
/note="4 bp duplication of IS6110 target sequence"

51..1142  
/gene="Rv0797"

51..1145  
/note="IS1547"

51..1145  
/gene="Rv0797"

51..1145  
/gene="Rv0797"

/note="Rv0797, (MTC1429B.03c-MTV042.07), transposase for  
IS1547, len 364 aa; almost identical to (but 20 aa shorter  
than) gpY13470(MTY13470\_2 Mycobacterium tuberculosis  
gene (383 aa). Also similar to other transposases eg.  
MIS110A.1 M:avium insertion sequence Q48909 transposase  
(464 aa); fasta scores: opt: 226 z-score: 275.7 E(1):  
2.4e-08, 30.7% identity in 199 aa overlap. Also slight  
similarity to M. tuberculosis protein MTCY39.03c (24.8%  
identity in 141 aa overlap)"

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv0797"

/protein\_id="CA809573.1"

/db\_xref="GI:3261812"

/db\_xref="SPTREMBL:O07182"

/translation="MVVGTDAHKTSHPFVATDEYGRUGKTYKATTAGATAYMA  
REQFGLLEWGEDCNMSARLEKLLAAGQVAVPPKLAQTKSARSKSDP  
ALAVARAVRETDPLATDETSREKLTDEADVAVQRTSAIRLMLVHEIDPER  
APARSLDAKHOALRTWLDTPQGLVLELAASLTDILRTGELNITLQISARVQ  
VALLEIFGCALTEAKITGSAAGTVRSAAACAAVAPIVMSGNAGOMRSL  
RSGNRDLNMLRIATQIRMPDSCGAAVYGRLODAGTKRAALCLRLRIAPVFOA  
LRVHPSSSHPTPPAACHRSYCSNCSLGS"

513..2334

/organism="Mycobacterium tuberculosis"

/strain="H37Rv"

/db\_xref="taxon:1773"

/clone="1429B"

/complement(1135..1932)

/gene="Rv0798c"

/complement(1135..1932)

/note="Rv0798c, (MTC1429B.02), len: 265 aa; 29 kDa  
antigen: cfp29, identical to gpY12820(MTY12820.1  
Mycobacterium tuberculosis cfp29 (265 aa), 99.6% identity  
in 265 aa overlap. Also highly similar Q45296 linocin M18  
from BL1NM18p\_1 B:linens; 58.5% identity in 265 aa  
overlap"

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv0798c"

/protein\_id="CA809572.1"

/db\_xref="GI:2190473"

/db\_xref="SPTREMBL:O07181"

/translation="MNNLYRDLPVTEAAAELELEAARTFRKHNGRRVVDSPGC  
PVTAASVGRGLIVKAPNTNGVIAHLASRKLIVRPLTSREIDVDYSGKSDME  
RVAKAARLAFVEDRTIFPEYSAASLEGIRASASNALPLPEPREIPVISOAISEL  
RLAACVDPSYVLASADYTKVSESHGIEIRHLNRIVDGDITARPADGAFVLTFR  
GGDDLDLGDVADYASHTDITVRLDTFLCYTAASVALSH"

/complement(11929..2936)  
/gene="Rv0799c"  
complement(1929..2936)  
/gene="Rv0799c"  
/note="Rv0799c, (MTCY0797A.10-MTC1429B.01), len: 335.  
similar to ECAB000330.8 E. coli K-12 P76536 (308 aa) E(1):  
2.2e-30, 37.4% identity in 297 aa overlap. Also similar to  
M. leprae, Q50021, U2266C, (146 aa), fasta scores, opt:  
147 z-score: 188.8 E(1): 0.0016, (33.3% identity in 117 aa  
overlap)and Q50020 U2266B, (27 aa), fasta scores, opt:  
94, E(1): 1.3, (56.5% identity in 23 aa overlap)"

```

/codon_start=1
/product="hypothetical protein Rv0799c"
/protein_id="CA09574.1"
/db_xref="GI:3261813"
/db_xref="SPTREMBL:007180"
/translation="MAVPAVSPOPILALPLPAALFIVATIGAGCEATVHDAISKISGL
VRAIGFRDPTRKLSVYVSGSDAMDRLFAPRPTLELPPELTIGPRHTAPATPGDILF
HIREPMDCPEELAGRLKSKMGDAVTVADHGRFPDNDLLGVDGTEPSPGLAI
KATTIGEDNRKFGSCYVHVKYVHDMASWESLVEQERIGRTKIDIDELDNKAP
ANSHVALNLTITDDGTERKIVRNHMPGEVKGEGYIFGYSRTPTVEOMLRNMF
GDPAGNTDRVLDLSTAVTGGLFSPITFDLHPPLPQAAITPLAGSLISGLKSKSP
R"
complement(1941..1945)
/gene="Rv0799c"
/feature="possible RBS upstream of Rv0798c"
2331..22980
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="y7H7A"
BASE COUNT      566 a      965 c      939 g      510 t
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 3; Length 2980;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 attcgatcctcgca 18
|||||
Db 2424 ATTTCGATCTCGTCA 2439

```

```

RESULT 18
MSGB13GS      42923 bp      DNA      BCT      15-JUN-1996
DEFINITION    Mycobacterium leprae cosmid B13 DNA sequence.
ACCESSION     L78823
VERSION       L78823.1 GI:1377768
KEYWORDS
SOURCE        Mycobacterium leprae (clone: cosmid B13) (tissue library: Lorient 6)
              DNA.
              Mycobacterium leprae
              Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
              Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
              Mycobacterium.
              1 (sites)
              Eiglmeyer, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.
              Use of an ordered cosmid library to deduce the genomic organization
              of Mycobacterium leprae
              Mol. Microbiol. 7 (2), 197-206 (1993)
              93188700
              2 (bases 1 to 42923)
              Smith, D.R., Richerich, P., Rubenfield, M., Butler, C., Lee, H.-M.,
              Xu, Q., Gunderson, K., Chung, M., Maher, J.K., Deloughery, C.,
              Aldrich, T., Smith, J., Tuil, C., Smyth, A., Drill, S., Avrich, A.S.,
              Ridge, P., Abendschan, K., Aldredge, T., Deloughery, C., Kirst, S.,
              Safer, H., Connolly, S., McDougall, S., Eiglmeyer, K., Bergh, S.,
              Cole, S., Robinson, K., Jaehn, L., Gryan, G., Church, G.M. and Mao, J.
              Prepublication submission
              Unpublished (1996)
              This sequence was generated by the Genome Sequencing Center at
              Genome Therapeutics Corporation (Collaborative Research Division),
              100
              Beaver St., Waltham, MA, 02154. Please contact Doug Smith
              (smith@ctc.com) for further information. The sequence represents
              the
              insert of a Lorient 6 cosmid clone from a mapped set of clones
              constructed from M. leprae genomic DNA isolated from armadillo
              liver
              [3]. The sequence may not represent the entire cloned insert of

```

cosmid if an overlapping region was previously sequenced from another clone. Coding sequences larger than 60 amino acids were predicted on the basis of codon usage and homology information. An attempt was made to locate the most probable start site based on codon usage, homology, the presence of a Shine-Dalgarno sequence, or overlapping of that suggested translational coupling. It is possible that the actual start site differs from the one selected.

```

FEATURES
  source
    1..42923
    /organism="Mycobacterium leprae"
    /specific_host="Dasyus novemcinctus"
    /db_xref="taxon:1769"
    /clone="cosmid B13"
    /tissue="liver of the armadillo"
    /note="The liver of the armadillo was used to isolate the
    Mycobacterium leprae."
BASE COUNT      9493 a      13528 c      11679 g      8223 t
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 3; Length 42923;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 ttcgatcctcgcaag 20
|||||
Db 20721 TTGATCTCGTCA 20736

```

```

RESULT 19
MCL1458/c
LOCUS         MCL1458      43839 bp      DNA      BCT      27-AUG-1999
DEFINITION    Mycobacterium leprae cosmid L458.
ACCESSION     AL049478
VERSION       AL049478.1 GI:4539121
KEYWORDS
SOURCE        Mycobacterium leprae.
              Mycobacterium leprae
              Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
              Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
              Mycobacterium.
              1 (bases 1 to 43839)
              Eiglmeyer, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.
              Use of an ordered cosmid library to deduce the genomic organization
              of Mycobacterium leprae
              Mol. Microbiol. 7 (2), 197-206 (1993)
              93188700
              2 (bases 1 to 43839)
              Seeger, K.J. and Harris, D.
              Unpublished
              3 (bases 1 to 43839)
              James, K.D., Parkhill, J., Barrell, B.G. and Randal, M.A.
              Direct Submission
              Submitted (15-MAR-1998) Mycobacterium leprae sequencing project,
              Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
              CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr.
              Stewart T. Cole. [3] Unite de Genetique Moleculaire Bacterienne,
              Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
              France Requests for cosmids should be sent to Karin Eiglmeyer
              (keiglmeyer@pasteur.fr)
              Notes:
              The Sanger Centre is funded to complete the sequence of M. leprae
              by the Helser Program for Research in Leprosy and Tuberculosis of

```

COMMENT

The New York Community Trust.  
 Work in Paris is supported by the Heiser Trust, the Association  
 Française Raoul Follereau and the Groupement de Recherches et des  
 Etudes des Genomes (GIP-GRG).  
 Details of M. lepreae sequencing at the Sanger Centre are available  
 on the World Wide Web.  
 (URL: <http://www.sanger.ac.uk/Projects/>)  
 CDS are numbered using the following system eg MLCB3.01c. ML (M.  
 lepreae), Cb33 (cosmid name), .01 (first CDS), c (complementary  
 strand).

The more significant matches with motifs in the PROSITE database  
 are also included but some of these may be fortuitous. The length  
 in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for  
 CDS which show significant similarity to other CDS in the database.  
 The position of possible ribosome binding site sequences are given  
 where these have been used to deduce the initiation codon. All CDS  
 over 100 codons have been analysed. Gene prediction is based on  
 positional base preference in codons especially where there is an  
 increase in the observed/expected third position G + C. CAUTION:  
 We may not have predicted the correct initiation codon. Where  
 possible we choose an initiation codon (atg, gtg, or ttg) which is  
 preceded by an upstream ribosome binding site sequence (optimally  
 5-13bp before the initiation codon). If this cannot be identified  
 we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the  
 sequenced clone. It may be shorter because we only sequence  
 overlapping sections once, or longer, because we arrange for a  
 small overlap between neighbouring submissions.

## FEATURES

## source

1. 43839

/organism="Mycobacterium lepreae"

/db\_xref="taxon:1769"

/clone="cosmid L458"

## misc\_feature

1. 569

/note="overlap with cosmid B1701 from 19932 to 20500"

## gene

complement(97..531)

/gene="MLCB485.01c"

## CDS

complement(97..531)

/gene="MLCB485.01c"

/note="MLCB485.01c, hypothetical protein, len: 114 aa;  
 unknown function, possible CDS based on amino acid  
 composition and frame analysis"

/codon\_start=1

/transl\_table=11

/label="MLCB485.01c"

/product="hypothetical protein MLCB485.01c"

/protein\_id="CAB39566.1"

/db\_xref="GI:4539122"

/translation="MOSLRHFTCRSTFEYHGYSIPAOAVYRTSTADTILDTMTGA  
 AGENORVYNQHPAVDQGRSLGSSVYVTRAPRSVGRTVSPADDLIDDSFCHSGSTNFFLD  
 OVAVFPSSADLKPOLKRTAKFRASVDGHEVQWLDVAVNIP"

complement(827..1210)

/gene="MLCB458.02c"

complement(827..1210)

/gene="MLCB458.02c"

/note="MLCB458.02c, hypothetical protein, len: 127 aa;  
 unknown function, N-terminus similar to internal sequence  
 of YP22\_MYCTU (EMBL:Z80226) RV0778 (MTCY369.22)

M.tuberculosis hypothetical protein (414 aa), fasta  
 scores: opt: 139 z-score: 189.2 E(): 0.0031, 31.4%

identity in 102 aa overlap"

/codon\_start=1

/transl\_table=11

/label="MLCB458.02c"

/product="hypothetical protein MLCB458.02c"

/protein\_id="CAB39567.1"

/db\_xref="GI:4539123"

/translation="MLEKNRPQIDFSDBILAKSPDEAMLSRRKIDLSHTFEYTLK  
 LIAEKRRNSDDIMRTLASAVITGNDKEFRILPANLEFFVLFTGNTAKHTTGYM  
 TAGVHEHSRPDKTIPRSGSIAPQR"

complement(1581..1850)

/gene="MLCB458.03c"

## CDS

complement(1581..1850)

/gene="MLCB458.03c"

/note="MLCB458.03c, hypothetical protein, len: 89 aa;  
 unknown function, improbable CDS based on frame analysis"

/codon\_start=1

/transl\_table=11

/label="MLCB458.03c"

/product="hypothetical protein MLCB458.03c"

/protein\_id="CAB39568.1"

/db\_xref="GI:4539124"

/translation="MQLAPRWGGEPEFTIAVKVYITILSHIONGRFSRIEYSL  
 FTRHRLICSSSDGGVGKSDPEPRRPSVRSNRTGSELP"

complement(1841..2110)

/gene="MLCB458.04c"

complement(1841..2110)

/gene="MLCB458.04c"

/note="MLCB458.04c, hypothetical protein, len: 89 aa;  
 unknown function, possible CDS based on amino acid  
 composition and frame analysis"

/codon\_start=1

/transl\_table=11

/label="MLCB458.04c"

/product="hypothetical protein MLCB458.04c"

/protein\_id="CAB39569.1"

/db\_xref="GI:4539125"

/translation="MTERMTAWRIPIPKIAREIPATALVSNPKIPLMPABSYRM  
 RCDPTSYRVDSFSDFYIDSHOSLSIYAVERSNOIKQLAISGCS"

3625..4344

/gene="MLCB458.05"

3625..4344

/gene="MLCB458.05"

/note="MLCB458.05, possible secreted protein, len: 239 aa;  
 unknown function, similar to TR:P95028 (EMBL:283863)

RV2525c (MTCY159.31) M.tuberculosis hypothetical protein  
 (240 aa), fasta scores: opt: 1358 z-score: 1535.9 E(): 0,  
 82.2% identity in 241 aa overlap. Contains possible  
 N-terminal signal sequence"

/codon\_start=1

/transl\_table=11

/label="MLCB458.05"

/product="hypothetical protein MLCB458.05"

/protein\_id="CAB39570.1"

/db\_xref="GI:4539126"

/translation="MSVSRPDLKFAVTVPGLIGLGVAAALCAVPASTAGSGTLTD  
 YAAVYTPASQIRATGAGATRYVSDPRGTANAGCPLOVETARDLNGIKTIVCYOG  
 KGNTRADWDGATAGLRHAQKGVOLHTAGGVSAPITASIDNPTTYOQYAOVAPTR  
 SMESVIGHQRTGVANSRTITAMLDGLASFWQHNMGSPKGYTHHPANLHOYEIDRR  
 TVGGVGVNVTILKPGFGWNA"

4956..14186

/gene="fas"

4956..14186

/gene="fas"

/note="MLCB458.06, fas, probable type I fatty acid  
 synthase, len: 3076 aa; similar to many eubacterial and  
 eukaryotic FAS e.g. TR:059497 (EMBL:X87822) Brevibacterium  
 ammoniagenes FAS (3063 aa), fasta scores: opt: 5361  
 z-score: 5527.0 E(): 0, 44.3% identity in 3106 aa overlap.

N-terminus shows similarity to FAS1-CANAL (EMBL:X74952)  
 Candida albicans FAS beta subunit (2037 aa) (27.3%  
 identity in 1928 aa overlap) and C-terminus to FAS2-CANAL  
 (EMBL:J29063) Candida albicans FAS alpha subunit (1885 aa)  
 (27.1% identity in 1739 aa overlap). C-terminus similar to  
 N-terminus of TR:069474 (EMBL:AL023635) MLCB1243.20c  
 kasA, M.lepreae beta-ketoacyl-ACP synthase (414 aa) (27.5%  
 identity in 386 aa overlap) and TR:069473 (EMBL:AL023635)  
 MLCB1243.19c, kasB, M.lepreae beta-ketoacyl-ACP synthase  
 (425 aa) (25.1% identity in 382 aa overlap). Equivalent to  
 TR:P95029 (EMBL:283863) RV2524c (MTCY159.32-MTY009.09c)  
 M.tuberculosis FAS (3069 aa) (85.8% identity in 3081 aa  
 overlap). Contains Pfam match to entry PF00698  
 Acyl-transf. Acyl transferase domain, score 18.20, E-value  
 1.9e-08, PS00017 ATP/GTP-binding site motif A (P-loop) and  
 PS00606 Beta-ketoacyl synthases active site"

/codon\_start=1  
/transl\_table=11  
/label=fas  
/product="putative type I fatty acid synthase"  
/protein\_id="CA839571.1"  
/db\_xref="GI:4539127"  
/translation="MTIHEHDQVSADRNGLSHGSRALADRLKAGEPVYVAFSGGSA  
WLETLLELVSAAGLEADLALVCEVLELLLEPVAKELVVPVPIGEPLQWRLALADL  
VPSDKHLTSAVSPGVLLQIAVGRALAQMDLITPVGIVGHSGVLAWEALKA  
GGARDALLMAOLIGAGTLVARRGIVSGDRPPVSVTNADPERIRLDEPAD  
VRVLPVLSIRNGRNSVITGTEPGLSREPRGROISDEEDRRKKIRGGDIPAPV  
FDPVVEIGFHTPLADGIGTGMMAKNGLDVTLAELTEALIVRQVVRITRYH  
GAGRYATIDGPDILIRIAPYRIGCVIVYANRGORTITVGAPEVVRAMIS  
YAPTVOLPGRILKSTKFTLRGSPFTLLAGMPTVDANIVAAANAGHMAELAG  
GVYETIFANRVEDLSGLEPGRYFNALDLPYLMKLOVGGKRIUOKRSGAID  
GVVTSGLDLELDELLELGGISVYVEKPTLEIQRVIRIATEMTRKPVIMHV  
EGRAGSHSMEIDLILLATYSELRSNANITVCGGIGTPEKAAEYLSGMAQAYG

Query Match 64.0% Score 16; DB 3; Length 43839;  
Best Local Similarity 100.0%; Pred. No. 11;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 31291 TTGCATCCTCGTCAGG 31276

RESULT 20

AC017911

LOCUS AC017911 118100 bp DNA HTG 09-DEC-1999  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
pieces.

AC017911

AC017911.1 GI:6553279

HTG; HTGS\_PHA52.

KEYWORDS fruit fly.

SOURCE Drosophila melanogaster

ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 118100).

AUTHORS Adams, M. and Venter, J. C.

TITLE Direct Submission

JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA

COMMENT This sequence was identified as CDW:10212606 by the submitter.  
For more information on this record e-mail to fly@celera.com.

\* NOTE: This is a working draft sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

FEATURES

source 1. 118100

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 32824 a 26059 c 26550 g 32667 t

ORIGIN

Query Match 64.0% Score 16; DB 65; Length 118100;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 63425 GCATTGCATCTCTGCT 63440

Query 1 gcatcgcacgcctcgt 16

|||||

AC007476 171705 bp DNA INV 17-MAR-2001

LOCUS AC007476 171705 bp DNA INV 17-MAR-2001

DEFINITION Drosophila melanogaster, chromosome 2R, region 49A-49B, BAC clone

BACR37P06, complete sequence.

AC007476 GI:13374640  
AC007476.5  
HTG.  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 171705)

AUTHORS

Celniker, S. E., Adams, M. D., Krommiller, B., Tyler, D., Wan, K. H.,  
Holt, R. A., Evans, C. A., Gockyne, J. D., Amanatides, P. G., Brandon, R. C.,  
Rogers, Y., An, H., Baldwin, D., Bancon, J., Beeson, K. Y., Busam, D. A.,  
Carlson, J. W., Center, A., Champagne, M., Davenport, L. B., Dietz, S. M.,  
Dodson, K., Dorsett, V., Doup, L. E., Doyle, C., Dresnek, D., Farfan, D.,  
Fertig, S., Frise, E., Galle, R. F., Gary, N. S., George, R. A.,  
Gonzalez, M., Houck, J., Hoskins, R. A., Hostin, D., Howland, T. J.,  
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,  
McIntosh, T. C., Moy, M., Murphy, B., Nelson, C., Nelson, K. A., Nunoo, J.,  
Paclob, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,  
Phouanavong, S., Pittman, G. S., Putli, V., Richards, S., Scheeler, F.,  
Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S. M.,  
Zaveri, J. S., Smith, H. O., Rubin, G. M., and Venter, J. C.

TITLE

JOURNAL

Unpublished

2 (bases 1 to 171705)

REFERENCE

AUTHORS

Celniker, S. E., Agbayani, A., Arcata, T. T., Baxter, E., Blazek, R. G.,  
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C. M., Farfan, D. E., Galle, R., George, R. A., Harris, N. L.,  
Hoskins, R. A., Houston, K. A., Hummasti, S. R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomtan, M. A., Mazda, P.,  
Moshrefi, A. R., Moshrefi, M., Nixon, K., Paclob, J. M., Park, S.,  
Pfeiffer, B., Poon, L., Segueira, A., Sethi, H., Shit, E.,  
Svirska, R. R., Wan, K. H., Weinburg, T., Zhang, R., Zieran, L. L. and  
Rubin, G. M.

TITLE

JOURNAL

Submitted (05-MAY-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT On Mar 17 2001 this sequence version replaced gi:5670614.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome  
shotgun and from subclones of this BAC and its neighboring clones.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

FEATURES

source 1. 171705

/organism="Drosophila melanogaster"

/strain="y: cn bw sp"

/db\_xref="taxon:7227"

/chromosome="2R"

/map="49A-49B"

/clone="BACR37P06 (D594)"

/clone.lib="RPCI-98 (Roswell Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
pBAC3.6)"

BASE COUNT 47654 a 38437 c 38157 g 47457 t

ORIGIN

Query Match 64.0% Score 16; DB 4; Length 171705;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 106124 GCATTGCATCTCTGCT 106109

Query 1 gcatcgcacgcctcgt 16

|||||

Db 106124 GCATTGCATCTCTGCT 106109

RESULT 22







sequences  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 728)  
AUTHORS Rubio, V.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-1996) Victor Rubio, Biocnologia Microbiolana,  
Centro Nacional de, Biocnologia (CSIC-UAM), Campus  
Cantoblanco-UAM, Madrid, 28049, Spain  
FEATURES  
Source  
1..728  
/organism="Thanatephorus cucumeris"  
/strain="2LR14"  
/specific\_host="Tulipa sp."  
/db\_xref="taxon:118239"  
/tissue\_type="mycelia"  
/note="Isolated from tulip, from The Netherlands"  
1..294  
/note="Internal transcribed spacer 1; ITS1"  
295..449  
/product="5.8S ribosomal RNA"  
450..728  
/note="Internal transcribed spacer 2; ITS2"  
BASE COUNT 196 a 153 c 145 g 234 t  
ORIGIN

Query Match 60.0%; Score 15; DB 15; Length 728;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 cctgtcaggtgcag 25  
|||||  
Db 20 CCTGCTCAGGTGCA 34

RESULT 25  
LOCUS R1CCYTAID 1801 bp DNA PLN 14-APR-2000  
DEFINITION Oryza sativa gene for cytoplasmic aldolase, complete cds,  
clone:ALDc-a.  
ACCESSION D13512  
VERSION D13512.1 GI:218156  
KEYWORDS cytoplasmic aldolase.  
SOURCE Oryza sativa DNA, clone:ALDc-a.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
REFERENCE  
AUTHORS Tsusumi, K., Kagaya, Y., Hidaka, S., Suzuki, J., Tokairin, Y.,  
Hirai, T., Hu, D., L., Ishikawa, K. and Ejiri, S.  
TITLE Structural analysis of the chloroplastic and cytoplasmic  
aldolase-encoding genes implicated the occurrence of multiple loci  
in rice  
JOURNAL Gene 141 (2), 215-220 (1994)  
MEDLINE 94215906  
REFERENCE 2 (bases 1 to 1801)  
TITLE Tsuchiya, T.  
AUTHORS Direct Submission  
JOURNAL Submitted (28-OCT-1992) to the DDBJ/EMBL/Genbank databases. Tohru  
Tsuchiya, Iwate University, Department of Agriculture, 3-18-8 Ueda,  
Morioka, Iwate 020, Japan (Tel:0196-23-5171(ex.2776),  
Fax:0196-25-7667)  
COMMENT Submitted (28-OCT-1992) to DDBJ by:  
Tohru Tsuchiya  
Institute for Cell Biology and Genetics  
Faculty of Agriculture  
Iwate University  
3-18-8 Ueda  
Morioka, Iwate 020  
Japan  
Phone: 0196-23-5171 x2776  
Fax: 0196-24-5084.

FEATURES  
Source  
Location/Qualifiers  
1..1801  
/organism="Oryza sativa"  
/db\_xref="taxon:4530"  
/clone="ALDc-a"  
41..49  
TATA signal  
prim\_transcript 80..1801  
exon 80..179  
/number=1  
CDS  
join(152..179, 615..1663)  
/codon\_start=1  
/product="cytoplasmic aldolase"  
/protein\_id="BAA02729.1"  
/db\_xref="GI:218157"  
/translation="MSAVCGKKYKDLIRNAAYICPGKGLADESTGTCKRRASIN  
VENVDNRRAPELLFCIPGALQYISGVIIDETLYQTKGKGPVDTLKEGALPGI  
KYDKETIEVAGDKETTTGGHDDLSQCAKTYEAGARAKRAVLTGPNPQSLADI  
LNAQGLACTAIIQDENGVLPIVEPEHLDVGDPHDIDRCAYVSEVYLAQYKALNEHVL  
LEGTLLKPMWTPGSDARKVAPEVIAETVRLQRTVPAPVAIVFLSGGSEEAFL  
NLNANNKLSAKKPMWLSFSGRALQOSTLAKMAGKTEWVEKARAFLVRCANSEATL  
GTCKGDAVLGEGAEHLHKVKY"  
180..614  
/number=1  
exon 615..1801  
/number=2  
BASE COUNT 418 a 521 c 439 g 423 t  
ORIGIN

Query Match 60.0%; Score 15; DB 14; Length 1801;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctctgcaggtgca 24  
|||||  
Db 1551 TCCTCCTCAGGTGCA 1565

RESULT 26  
LOCUS AB037816 4886 bp mRNA PRI 14-MAR-2000  
DEFINITION Homo sapiens mRNA for KIAA1395 protein, partial cds.  
ACCESSION AB037816  
VERSION AB037816.1 GI:7243170  
KEYWORDS  
SOURCE Homo sapiens brain cDNA to mRNA, clone\_1lb:pb1uescriptII SK plus  
clone:hj07927.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Nagase, T., Kikuno, R., Ishikawa, K.I., Hirose, M. and Ohara, O.  
TITLE Prediction of the coding sequences of unidentified human genes.  
XVI. The complete sequences of 150 new cDNA clones from brain which  
code for large proteins in vitro  
JOURNAL DNA Res. 7 (1), 65-73 (2000)  
MEDLINE 20181126  
REFERENCE 2 (bases 1 to 4886)  
TITLE Ohara, O., Nagase, T. and Kikuno, R.  
AUTHORS Direct Submission  
JOURNAL Submitted (31-JAN-2000) to the DDBJ/EMBL/Genbank databases. Osamu  
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology,  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:odnainfo@kazusa.or.jp, URL: http://www.kazusa.or.jp/huge/,  
Tel: +81-438-52-3913, Fax: +81-438-52-3914)  
FEATURES  
Source  
Location/Qualifiers  
1..4886  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="hj07927"  
/clone\_1lb="pb1uescriptII SK plus"  
/tissue\_type="brain"  
1..4886  
gene





gene  
CDS  
complement(3254..4033)  
/gene="PA2811"  
/codon\_start=1  
/transl\_table=11  
/product="probable permease of ABC-2 transporter"  
/protein\_id="AAC06199.1"  
/db\_xref="GI:9948895"  
/translation="MSSELSANVALNTIYRREVRTRTIPOTLLPPIATWLYYVI  
EGLNIGQIDMGFTTYMETVPEGLIMSVITAYGVSSFFGSKRORSVEBLVSP  
VSPRTILGTTIGVGRGLAVGLVLMLEFTEGLIOVHHLGIVLVLTATFESLGG  
FINAVFARNFDISITPTFVLPTLVGLGVYSINLLPFWQTVSLANPLIHWVNAFR  
YGLIGVSDIRIGVAGFMFLATVLYIGCVRLVSGRMQ"  
/gene="PA2812"  
complement(4030..4962)  
/gene="PA2812"  
/codon\_start=1  
/transl\_table=11  
/product="probable ATP-binding component of ABC  
transporter"  
/protein\_id="AAC06200.1"  
/db\_xref="GI:9948896"  
/translation="MSSALSIQQLTKTYGNGQALKGIDVVAEGDFALLGNGAGK  
STTIGILSTLVNKTSGSVFEGHLDKDPAGLKGCVVPOEFNFQFKVDIYVVO  
AGYGIIPAKIAKRAERYLTOLGMDKRNDSFGLSGMKRRLMLRAIWHOPRLIIL  
DEPTAGVDIELRSRMSFTELNQEGISILTHYLEEAGQCRNTAIIIDHEIYONT  
SMRDLMTLSHTEFLDLKNVQALPTLGGYPTRLVDHTLVEVKSQINDPRAQL  
GAQIEVLSIRKNTNLEELFVSIVKNIIRIAR"  
complement(5038..5658)  
/gene="PA2813"  
complement(5038..5658)  
/gene="PA2813"  
/codon\_start=1  
/transl\_table=11  
/product="probable glutathione S-transferase"  
/protein\_id="AAC06201.1"  
/db\_xref="GI:9948897"  
/translation="MLTIWGRKNSNVKRAKLCWCAEAGLEYRTVDAGAGFGLVDEPAF  
RAMNPGRVVIEDGEFVLENSAIVRYIAAAYAGDILPODPVRAADKMDMTTS  
TLAGEPRDLFWGTLRTPRODEALAKLOLCGELLRRPDTLAGQPMLSERGMG  
DIPGCFIYAFEMPIERPPLHLAWYILRERPAFRVAVTELT"  
complement(5773..6444)  
/gene="PA2814"  
complement(5773..6444)  
/gene="PA2814"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC06202.1"  
/db\_xref="GI:9948898"  
/translation="MOSYLOSGRFVDSHPVVEFAEKSRSNGSAKPRDOAVAIYYAR  
LGVRYNPVYFSRDPOTILKASHALOGESECVKALLAACARHICIPRIGIAYVRI  
LAPRLLELRESEVPAMGVTETLEYEGWVATPFPNALCRCPVALDEPDVADVS  
FHPNRGGRMEVYLADHGCFADLPPELFFSHLOQHYHLSGRIALDGPQADAGQ  
DEGRN"  
6597..9044  
/gene="PA2815"  
6597..9044  
/gene="PA2815"  
/codon\_start=1  
/transl\_table=11  
/product="probable acyl-CoA dehydrogenase"  
/protein\_id="AAC06203.1"  
/db\_xref="GI:9948899"  
/translation="MLLMLVVLVGVAYLAHRRTPPAALGISAAVILIMGVSHAP  
GMLLVFWMLMAVALPALPDLRRVYSGPLFAMFOKVIPLPMDTEFEALAEACTVM  
DGELESGPDMOKLIDYPRKQLTREEDAFVNDGPELCAWSDMDIGORDLPEEANA  
YIKHGERGLITPEYGGKSAHSAHSYVKAITRSGDASTVWVNSGAPALLH  
YCTDORQRYITSLAKGDIPTCFALDIPYASDAGMTDVIVKCGWEGREYVGLL  
TWERYITLGFVALILGLAFKCHDPDLHLDGEDLIGITLALIPDTGVEGERHVL  
GAAGRNNGNSGKDVPLEYLIIGOEMIGKMMMLNCLISVGRISLIPAVGTSGAKS  
SVSGRYAOVREQFNVPILAEFEGIOEALARIGGNAMLMDSARILITANAVIDGEPSYL  
SAILKYLTERGRCIAHAMDIHGKGIIMGPNNYIGRSNOGAPITITVGCANILSN

gene  
CDS  
LMIFGGATRCHPYVLEKEMELAQREDKQDAEFDPALLVKHIGFAVNAASSPLTSL  
CHLGNAPGDRISRPYFRLALNLRASPLLDLDFSWMLLGGELKKRERISARIGVLSY  
LYGSAALKRHYHDLGNPDYIHLPLMAWEEENIGKREALEDLINPFRRFCALKYL  
VLPGRRHKGPCDELDLAEITLGRPDPPALQALILACAFLEKDPDQVGLAIAFPA  
VESASALEKTLKRAIKREGKVPQAGNPLDGAAGVLSABOQALHQAFAARRKVID  
VDVFAKEELQADGRIH"  
9173..9553  
/gene="PA2816"  
9173..9553  
/gene="PA2816"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC06204.1"  
/db\_xref="GI:9948900"  
/translation="MLKLFATGIAAFLLPAPAAAPAPYGVPHOAVORAGEQRO  
ROLQROQRFPBQRQLQDQDLOKQOQMLQROKROKQMDNLIRQQLDQQRWLEED  
ORLDSERQLENRRROSOPAIR"  
9612..10022  
/gene="PA2817"  
9612..10022  
/gene="PA2817"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC06205.1"  
/db\_xref="GI:9948901"  
/translation="MANSYLEHIALLOHLRNITGALGEAEOVPEENHGLFLEPDL  
MDELPRDEGAOYLGODLISOVPHRYQIATLVPRDLMEFGGDCLEFMPDELDLFO  
RLDERRYAEERGCFFPMNRKQALALPDSFRH"  
complement(10089..11666)  
/gene="PA2818"  
complement(10089..11666)

Query Match 60.0%; Score 15; DB 1; Length 12542;  
Best local similarity 100.0%; Pred. No. 49;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 cgatcctcgcagt 21  
|||||  
Db 4038 CGATCCTCGTCAGGT 4052

RESUBMIT 28  
AL499627  
LOCUS  
DEFINITION  
Human DNA sequence from clone RP13-379024 on chromosome 20 Contains  
ESTs, STSS, GSSs and two CpG islands, complete sequence.  
ACCESSION  
AL499627.23 GI:13559068  
KEYWORDS  
HTG; CpG island; GATA-5; transcription factor.  
SOURCE  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 59999)  
Mathews, L.  
Direct Submission  
Submitted (02-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquey@sanger.ac.uk  
On Apr 5, 2001 this sequence version replaced gi:13446506.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats: all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, SW: SWISSPROT, Tr: TREMBL, Wp: WORMPEP; information on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr20>  
 RP13-379024 is from the library RP13-13.2 constructed by the group of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBAC3.6  
 IMPORTANT: This sequence is not the entire insert of clone RP13-379024. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true right end of clone RP13-379024 is at 59939 in this sequence. The true right end of clone RP5-908M14 is at 100 in this sequence.

## FEATURES

Source 1..59939  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /clone="RP13-379024"  
 /clone\_lib="RP13-13.2"  
 /complement(1..95)  
 /note="match: STS: Em:HS908M14T"  
 1020..1119  
 /note="2 copies 50 mer 91% conserved"  
 1026..1121  
 /note="4 copies 24 mer 79% conserved"  
 3539..3867  
 /note="AluIo repeat: matches 3..306 of consensus"  
 3939..4116  
 /note="AluIo repeat: matches 54..220 of consensus"  
 5954..6007  
 /note="2 copies 27 mer 96% conserved"  
 6334..6456  
 /note="MER5A repeat: matches 1..125 of consensus"  
 6931..7062  
 /note="4 copies 33 mer 84% conserved"  
 /complement(7115)  
 /gene="B379024.1"  
 10044..10169,17021..17196,18617..>19251)  
 /gene="B379024.1"  
 /note="match: cDNAs: Em:U04725  
 match: ESTs: Em:H80372 Em:H80385 Em:W00677 Em:A1632657  
 Em:N72525 Em:H80289 Em:AA127875 Em:AA127913 Em:AW70386  
 Em:AA972923 Em:AF685293 Em:AA534131 Em:AW973240  
 Em:AA533669 Em:AA127923 Em:NA53566 Em:AW511251 Em:AI174441"  
 /evidence="not\_experimental"  
 /product="B379024.1 (novel protein similar to  
 transcription factor GATA-5)"  
 complement(7115..19251)  
 /gene="B379024.1"  
 7724..8104  
 /note="match: STS: Em:G25060"  
 7724..8137  
 /note="match: STS: Em:G29730"  
 7764..7900  
 /note="MER1B repeat: matches 1..124 of consensus"  
 complement(join(8454..8609,8958..9082,9452..9539,  
 10044..10169,17021..17196,18617..19139))  
 /gene="B379024.1"  
 /note="match: proteins: Tr:Q16365 Sw:P43692 Wp:F52C12  
 Tr:Q9TXV3 Sw:P19212 Sw:P23772 Sw:P23825 Sw:Q91428

## misc\_feature

repeat\_region  
 10311..10434  
 /note="MIR repeat: matches 62..188 of consensus"  
 11461..11560  
 /note="2 copies 50 mer 86% conserved"  
 11596..11651  
 /note="2 copies 28 mer 94% conserved"  
 11672..11727  
 /note="2 copies 28 mer 94% conserved"  
 11744..11799  
 /note="2 copies 28 mer 94% conserved"  
 11820..11875  
 /note="2 copies 28 mer 96% conserved"  
 11892..11947  
 /note="2 copies 28 mer 96% conserved"  
 12774..12958  
 /note="MER53 repeat: matches 10..189 of consensus"  
 14602..14773  
 /note="MIR repeat: matches 37..214 of consensus"  
 15051..15342  
 /note="AluIo repeat: matches 7..299 of consensus"  
 15351..15663  
 /note="AluIo repeat: matches 1..296 of consensus"  
 17210..17291  
 /note="2 copies 41 mer 90% conserved"  
 17861..20459  
 /note="CpG island"  
 /evidence="not\_experimental"  
 19352..20175  
 /note="Single clone region. Sequence from reads from a  
 short insert library derived from a single pUC clone.  
 Restriction digest data confirm the assembly."  
 19448..19645  
 /note="match: GSS: Em:AQ938920"  
 19641..19712  
 /note="3 copies 24 mer 84% conserved"  
 19642..19713  
 /note="36 copies 2 mer CC 70% conserved"  
 complement(20394..20581)  
 /note="match: STS: Em:J30262"  
 20489..20956  
 /note="117 copies 4 mer ttcc 66% conserved"  
 20492..20722  
 /note="77 copies 33 mer 67% conserved"  
 20496..20959  
 /note="29 copies 16 mer 65% conserved"  
 20500..20950  
 /note="11 copies 41 mer 62% conserved"  
 20504..20951  
 /note="16 copies 28 mer 66% conserved"  
 20724..20789  
 /note="2 copies 33 mer 93% conserved"  
 20827..20925  
 /note="3 copies 33 mer 76% conserved"

```

repeat_region      20831..20920
                    /note="15 copies 6 mer ttttc 678 conserved"
misc_feature        20933
                    /note="Tandem repeat. Forced join. Gap size estimated to
                    be approximately 110bp by restriction digest data."
repeat_region      20958..21234
                    /note="Alusx repeat: matches 9..294 of consensus"
repeat_region      22223..22386
                    /note="MIR repeat: matches 52..226 of consensus"
repeat_region      22934..23051
                    /note="L2 repeat: matches 2349..2473 of consensus"
repeat_region      23253..23314
                    /note="MIR repeat: matches 122..183 of consensus"
repeat_region      23676..23871
                    /note="7 copies 28 mer 628 conserved"
repeat_region      23687..23851
                    /note="5 copies 33 mer 848 conserved"
repeat_region      23689..23877
                    /note="7 copies 27 mer 648 conserved"
repeat_region      23697..23846
                    /note="3 copies 50 mer 748 conserved"
repeat_region      23741..23868
                    /note="8 copies 16 mer 678 conserved"
repeat_region      25184..25344
                    /note="MIR repeat: matches 21..190 of consensus"
repeat_region      27265..27342

Query Match
Best Local Similarity 100.0%; Pred. No. 48;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      9 atccctgcaggtgc 23
        |||||||
Db      12634 ATCCTGCTAGGTGC 12648

RESULT 29
AP003452
LOCUS      AP003452 168972 bp DNA
DEFINITION Oryza sativa chromosome 1 clone P0478H03, *** SEQUENCING IN
PROGRESS *** in ordered pieces.
ACCESSION AP003452
VERSION    AP003452.1 GI:13486921
KEYWORDS   HTG: HTGS_PHASE2
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0478H03.
ORGANISM    Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
            Oryza.
REFERENCE   1 (sites)
AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLES     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0478H03
JOURNAL    Published Only in Database (2001) In press
2 (bases 1 to 168972)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (28-MAR-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@agr.affrc.go.jp, URL:http://rgrp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
            location/Qualifiers
FEATURES
JOURNAL
AUTHORS
TITLES
REFERENCE
COMMENT

```

## FEATURES

```

source
1..168972
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0478H03"
BASE COUNT 46333 a 37251 c 37299 g 47789 t 300 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 47;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      10 tccctgcaggtgca 24
        |||||||
Db      165052 TCCTGCTAGGTGCA 165066

RESULT 30
AP003455/c
LOCUS      AP003455/c 193530 bp DNA
DEFINITION Oryza sativa chromosome 1 clone P0519D04, *** SEQUENCING IN
PROGRESS *** in ordered pieces.
ACCESSION AP003455
VERSION    AP003455.1 GI:13486924
KEYWORDS   HTG: HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0519D04.
ORGANISM    Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
            Oryza.
REFERENCE   1 (sites)
AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLES     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0519D04
JOURNAL    Published Only in Database (2001) In press
2 (bases 1 to 193530)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (28-MAR-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@agr.affrc.go.jp, URL:http://rgrp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
            location/Qualifiers
FEATURES
source
1..193530
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0519D04"
BASE COUNT 59449 a 41978 c 42335 g 54218 t 50 others
ORIGIN

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 47;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      10 tccctgcaggtgca 24
        |||||||
Db      148848 TCCTGCTAGGTGCA 148834

```

RESULT 31  
AF140181/c 483 bp DNA PLN 12-FEB-2001  
LOCUS Filobasidiella neoformans isolate CN19.95 diphenol oxidase gene,  
DEFINITION partial cds.  
ACCESSION AF140181  
VERSION AF140181.1 GI:5734019  
KEYWORDS  
SOURCE Filobasidiella neoformans.  
ORGANISM Filobasidiella neoformans.  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;  
Tremellales; Tremellaceae; Filobasidiella.  
REFERENCE 1 (bases 1 to 483)  
AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.  
TITLE Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*  
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)  
MEDLINE 20504730  
PUBMED 11050543  
REFERENCE 2 (bases 1 to 483)  
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.  
TITLE Direct Submission  
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke University, Research Dr., Durham, NC 27710, USA  
FEATURES  
source  
1..483  
/organism="Filobasidiella neoformans"  
/isolate="CN19.95"  
/db\_xref="taxon:5207"  
join(<75..134,187..302,355..>442)  
/product="diphenol oxidase"  
join(<75..134,187..302,355..442)  
/note="laccase precursor"  
/codon\_start=1  
/product="diphenol oxidase"  
/protein\_id="AAD49841.1"  
/db\_xref="GI:5734053"  
/translation="MHGMKRNTPYMDGVPGITCPIPPGSGSTYNTISDSGTYWM  
HSHSNMADGLMGPILVHSVDEPIORGRDDEDRIVVTWMH"  
BASE COUNT 124 a 108 c 108 g 143 t  
ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 483;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcgtcctcgta 18  
|||||  
Db 416 TTCGATCCTCGTCA 403

RESULT 32  
AF140179/c 509 bp DNA PLN 12-FEB-2001  
LOCUS Filobasidiella neoformans isolate CN-C diphenol oxidase gene,  
DEFINITION partial cds.  
ACCESSION AF140179  
VERSION AF140179.1 GI:5734017  
KEYWORDS  
SOURCE Filobasidiella neoformans.  
ORGANISM Filobasidiella neoformans.  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;  
Tremellales; Tremellaceae; Filobasidiella.  
REFERENCE 1 (bases 1 to 509)  
AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.  
TITLE Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*  
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)

MEDLINE 20504730  
PUBMED 11050543  
REFERENCE 2 (bases 1 to 509)  
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.  
TITLE Direct Submission  
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke University, Research Dr., Durham, NC 27710, USA  
FEATURES  
source  
1..509  
/organism="Filobasidiella neoformans"  
/isolate="CN-C"  
/db\_xref="taxon:5207"  
join(<86..145,198..313,366..449,503..>509)  
/product="diphenol oxidase"  
join(<86..145,198..313,366..449,503..>509)  
/note="laccase precursor"  
/codon\_start=1  
/product="diphenol oxidase"  
/protein\_id="AAD49839.1"  
/db\_xref="GI:5734051"  
/translation="MHGMKRNTPYMDGVPGITCPIPPGSGSTYNTISDSGTYWM  
HSHSNMADGLMGPILVHSVDEPIORGRDDEDRIVVTWMH"  
BASE COUNT 135 a 108 c 121 g 145 t  
ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 509;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcgtcctcgta 18  
|||||  
Db 427 TTCGATCCTCGTCA 414

RESULT 33  
AF140176/c 513 bp DNA PLN 12-FEB-2001  
LOCUS Filobasidiella neoformans isolate ZG284 diphenol oxidase gene,  
DEFINITION partial cds.  
ACCESSION AF140176  
VERSION AF140176.1 GI:5734014  
KEYWORDS  
SOURCE Filobasidiella neoformans.  
ORGANISM Filobasidiella neoformans.  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;  
Tremellales; Tremellaceae; Filobasidiella.  
REFERENCE 1 (bases 1 to 513)  
AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.  
TITLE Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*  
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)  
MEDLINE 20504730  
PUBMED 11050543  
REFERENCE 2 (bases 1 to 513)  
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.  
TITLE Direct Submission  
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke University, Research Dr., Durham, NC 27710, USA  
FEATURES  
source  
1..513  
/organism="Filobasidiella neoformans"  
/isolate="ZG284"  
/db\_xref="taxon:5207"  
join(<76..135,188..303,356..439,493..>513)  
/product="diphenol oxidase"  
join(<76..135,188..303,356..439,493..>513)  
/note="laccase precursor"  
/codon\_start=1  
/product="diphenol oxidase"  
/protein\_id="AAD49836.1"  
/db\_xref="GI:5734048"

```

/translation="MHGMROKNSPYMDGIPGTCPIPPGSGTYTNYFTISDQSGTYWM
BASE COUNT      136 a      115 c      116 g      146 t
ORIGIN

Query Match      56.0%; Score 14; DB 13; Length 513;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 ttgcgtccctgc tca 18
        |||
Db      417 TTGCATCTCGTCA 404

RESULT  34
AF140173      526 bp      DNA      PLN      12-FEB-2001
LOCUS      Filobasidiella neoformans isolate MMRL750 diphenol oxidase gene,
DEFINITION      partial cds.
ACCESSION      AF140173
VERSION      AF140173.1 GI:5734011
KEYWORDS
SOURCE
ORGANISM      Filobasidiella neoformans.
               Filobasidiella neoformans.
               Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
               Tremellales; Tremellaceae; Filobasidiella.
REFERENCE      1 (bases 1 to 526)
AUTHORS      Xu,J., Vilgalys,R. and Mitchell,T.G.
TITLE      Multiple gene genealogies reveal recent dispersion and
               hybridization in the human pathogenic fungus Cryptococcus
               neoformans
JOURNAL      Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE      20504730
PUBMED      11050543
REFERENCE      2 (bases 1 to 526)
AUTHORS      Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE      Direct Submission
JOURNAL      Submitted (01-APR-1999) Department of Microbiology, Duke
               University, Research Dr., Durham, NC 27710, USA
FEATURES
source
1..526
/organism="Filobasidiella neoformans"
/isolate="MMRL750"
/db_xref="taxon:5207"
join(<96..155,206..321,376..459,512..>526)
/product="diphenol oxidase"
join(<96..155,206..321,376..459,512..>526)
/feature="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD4983.1"
/db_xref="GI:5734045"
/translation="MHGLRQLGTAFMGVGATGTCPIPPGSGTYTNYFTISDQSGTYWM
               HSHYSNMAADGLMGPLIVHSVHPIQGRDVEDRIVFTVMHMDNSEI"
BASE COUNT      144 a      121 c      116 g      145 t
ORIGIN

Query Match      56.0%; Score 14; DB 13; Length 526;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 ttgcgtccctgc tca 18
        |||
Db      437 TTGCATCTCGTCA 424

RESULT  35
AF140183      528 bp      DNA      PLN      12-FEB-2001
LOCUS      Filobasidiella neoformans isolate CN140.97 diphenol oxidase gene,
DEFINITION      partial cds.

```

```

ACCESSION      AF140183
VERSION      AF140183.1 GI:5734021
KEYWORDS
SOURCE
ORGANISM      Filobasidiella neoformans.
               Filobasidiella neoformans.
               Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
               Tremellales; Tremellaceae; Filobasidiella.
REFERENCE      1 (bases 1 to 528)
AUTHORS      Xu,J., Vilgalys,R. and Mitchell,T.G.
TITLE      Multiple gene genealogies reveal recent dispersion and
               hybridization in the human pathogenic fungus Cryptococcus
               neoformans
JOURNAL      Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE      20504730
PUBMED      11050543
REFERENCE      2 (bases 1 to 528)
AUTHORS      Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE      Direct Submission
JOURNAL      Submitted (01-APR-1999) Department of Microbiology, Duke
               University, Research Dr., Durham, NC 27710, USA
FEATURES
source
1..528
/organism="Filobasidiella neoformans"
/isolate="CN140.97"
/db_xref="taxon:5207"
join(<96..155,208..323,376..459,513..>528)
/product="diphenol oxidase"
join(<96..155,208..323,376..459,513..>528)
/feature="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD4984.1"
/db_xref="GI:5734055"
/translation="MHGMROKNTPYMDGIPGTCPIPPGSGTYTNYFTISDQSGTYWM
               HSHYSNMAADGLMGPLIVHSVHPIQGRDVEDRIVFTVMHMDNSEI"
BASE COUNT      140 a      115 c      123 g      150 t
ORIGIN

Query Match      56.0%; Score 14; DB 13; Length 528;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 ttgcgtccctgc tca 18
        |||
Db      437 TTGCATCTCGTCA 424

RESULT  36
AF140164      529 bp      DNA      PLN      12-FEB-2001
LOCUS      Filobasidiella neoformans isolate CN124.91 diphenol oxidase gene,
DEFINITION      partial cds.
ACCESSION      AF140164
VERSION      AF140164.1 GI:5734002
KEYWORDS
SOURCE
ORGANISM      Filobasidiella neoformans.
               Filobasidiella neoformans.
               Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
               Tremellales; Tremellaceae; Filobasidiella.
REFERENCE      1 (bases 1 to 529)
AUTHORS      Xu,J., Vilgalys,R. and Mitchell,T.G.
TITLE      Multiple gene genealogies reveal recent dispersion and
               hybridization in the human pathogenic fungus Cryptococcus
               neoformans
JOURNAL      Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE      20504730
PUBMED      11050543
REFERENCE      2 (bases 1 to 529)
AUTHORS      Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE      Direct Submission
JOURNAL      Submitted (01-APR-1999) Department of Microbiology, Duke
               University, Research Dr., Durham, NC 27710, USA

```

FEATURES  
Source Location/Qualifiers  
1..529  
/organism="Filobasidiella neoformans"  
/isolate="CN124.91"  
/db\_xref="taxon:5207"  
join(<93..152,203..318,373..456,509..>529)  
/product="diphenol oxidase"  
join(<93..152,203..318,373..456,509..>529)  
/note="laccase precursor"  
/codon\_start=1  
/product="diphenol oxidase"  
/protein\_id="AAD49824.1"  
/db\_xref="GI:5734036"  
/translation="WHGLRQLGTAFMGVPGITQCP1PPGSSFTYQFTVSHSGTFMW  
HSHYSNMGADGIMGPLIHSNPEPLRGDRYDEDRIVFTIDWVHNSSEVY"  
BASE COUNT 144 a 121 c 118 g 146 t  
ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 529;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcattccgtca 18  
|||||  
Db 434 TTCGATCCTCGTCA 421

RESULT 37  
AF140165/c 530 bp DNA PLN 12-FEB-2001  
LOCUS  
DEFINITION Filobasidiella neoformans isolate CN2109.91 diphenol oxidase gene,  
partial cds.  
AF140165  
ACCESSION  
VERSION AF140165.1 GI:5734003  
KEYWORDS  
SOURCE  
ORGANISM  
. Filobasidiella neoformans.  
Filobasidiella neoformans.  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;  
Tremellales; Tremellaceae; Filobasidiella.  
XU,J., Vilgalys,R. and Mitchell,T.G.  
1 (bases 1 to 530)  
Multiple gene genealogies reveal recent dispersion and  
hybridization in the human pathogenic fungus *Cryptococcus*  
*neoformans*  
Mol. Ecol. 9 (10), 1471-1481 (2000)  
JOURNAL  
MEDLINE 20504730  
PUBMED 11050543  
REFERENCE 2 (bases 1 to 530)  
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.  
TITLE Direct Submission  
SUBMITTED (01-APR-1999) Department of Microbiology, Duke  
University, Research Dr., Durham, NC 27710, USA  
LOCATION/Qualifiers  
1..530  
/organism="Filobasidiella neoformans"  
/isolate="CN2109.91"  
/db\_xref="taxon:5207"  
join(<94..153,204..319,374..457,510..>530)  
/product="diphenol oxidase"  
join(<94..153,204..319,374..457,510..>530)  
/note="laccase precursor"  
/codon\_start=1  
/product="diphenol oxidase"  
/protein\_id="AAD49825.1"  
/db\_xref="GI:5734037"  
/translation="WHGLRQLGTAFMGVPGITQCP1PPGSSFTYQFTVSHSGTFMW  
HSHYSNMGADGIMGPLIHSNPEPLRGDRYDEDRIVFTIDWVHNSSEVY"  
BASE COUNT 144 a 122 c 118 g 146 t  
ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 530;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcattccgtca 18  
|||||  
Db 435 TTCGATCCTCGTCA 422

RESULT 38  
AF140169/c 530 bp DNA PLN 12-FEB-2001  
LOCUS  
DEFINITION Filobasidiella neoformans isolate J10 diphenol oxidase gene,  
partial cds.  
AF140169  
ACCESSION  
VERSION AF140169.1 GI:5734007  
KEYWORDS  
SOURCE  
ORGANISM  
. Filobasidiella neoformans.  
Filobasidiella neoformans.  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;  
Tremellales; Tremellaceae; Filobasidiella.  
XU,J., Vilgalys,R. and Mitchell,T.G.  
1 (bases 1 to 530)  
Multiple gene genealogies reveal recent dispersion and  
hybridization in the human pathogenic fungus *Cryptococcus*  
*neoformans*  
Mol. Ecol. 9 (10), 1471-1481 (2000)  
JOURNAL  
MEDLINE 20504730  
PUBMED 11050543  
REFERENCE 2 (bases 1 to 530)  
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.  
TITLE Direct Submission  
SUBMITTED (01-APR-1999) Department of Microbiology, Duke  
University, Research Dr., Durham, NC 27710, USA  
LOCATION/Qualifiers  
1..530  
/organism="Filobasidiella neoformans"  
/isolate="J10"  
/db\_xref="taxon:5207"  
join(<94..153,204..319,374..457,510..>530)  
/product="diphenol oxidase"  
join(<94..153,204..319,374..457,510..>530)  
/note="laccase precursor"  
/codon\_start=1  
/product="diphenol oxidase"  
/protein\_id="AAD49829.1"  
/db\_xref="GI:5734041"  
/translation="WHGLRQLGTAFMGVPGITQCP1PPGSSFTYQFTVSHSGTFMW  
HSHYSNMGADGIMGPLIHSNPEPLRGDRYDEDRIVFTIDWVHNSSEVY"  
BASE COUNT 144 a 122 c 119 g 145 t  
ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 530;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcattccgtca 18  
|||||  
Db 435 TTCGATCCTCGTCA 422

RESULT 39  
AF140172/c 530 bp DNA PLN 12-FEB-2001  
LOCUS  
DEFINITION Filobasidiella neoformans isolate M0061 diphenol oxidase gene,  
partial cds.  
AF140172  
ACCESSION  
VERSION AF140172.1 GI:5734010  
KEYWORDS  
SOURCE  
ORGANISM  
. Filobasidiella neoformans.  
Filobasidiella neoformans.  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;  
Tremellales; Tremellaceae; Filobasidiella.

```

REFERENCE 1 (bases 1 to 530)
AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.
TITLE Multiple gene genealogies reveal recent dispersion and
hybridization in the human pathogenic fungus Cryptococcus
neoformans
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE 20504730
PUBMED 11050543
REFERENCE 2 (bases 1 to 530)
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
University, Research Dr., Durham, NC 27710, USA
FEATURES
Source Location/Qualifiers
1..530
/organism="Filobasidiella neoformans"
/isolate="M0061"
/db_xref="taxon:5207"
join(<94..153,204..319,374..457,510..>530)
/product="diphenol oxidase"
join(<94..153,204..319,374..457,510..>530)
/note="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD49832.1"
/db_xref="GI:5734044"
/translation="MHGLROLGTAFMGVPGITQCPPIPGSSFTYQFTVSHSGTFMW
HSHYSNMAIGIMGPLITHSPNEPLQGRYDDEDRIVFTIDWHDNSEVY"
BASE COUNT 144 a 123 c 118 g 145 t
ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcgtcctcgta 18
|||||
Db 435 TTCGATCCTCGTCA 422

RESULT 40
AF140175 530 bp DNA PLN 12-FEB-2001
LOCUS Filobasidiella neoformans isolate ZG280 diphenol oxidase gene,
DEFINITION partial cds.
ACCESSION AF140175
VERSION AF140175.1 GI:5734013
KEYWORDS
SOURCE
ORGANISM Filobasidiella neoformans.
Filobasidiella neoformans
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
Tremellales; Tremellaceae; Filobasidiella.
1 (bases 1 to 530)
Xu,J., Vilgalys,R. and Mitchell,T.G.
Multiple gene genealogies reveal recent dispersion and
hybridization in the human pathogenic fungus Cryptococcus
neoformans
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE 20504730
PUBMED 11050543
REFERENCE 2 (bases 1 to 530)
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
University, Research Dr., Durham, NC 27710, USA
FEATURES
Source Location/Qualifiers
1..530
/organism="Filobasidiella neoformans"
/isolate="ZG280"
/db_xref="taxon:5207"
join(<94..153,204..319,374..457,510..>530)
/product="diphenol oxidase"

```

```

CDS
join(<94..153,204..319,374..457,510..>530)
/note="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD49835.1"
/db_xref="GI:5734047"
/translation="MHGLROLGTAFMGVPGITQCPPIPGSSFTYQFTVSHSGTFMW
HSHYSNMAIGIMGPLITHSPNEPLQGRYDDEDRIVFTIDWHDNSEVY"
BASE COUNT 144 a 122 c 118 g 146 t
ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcgtcctcgta 18
|||||
Db 435 TTCGATCCTCGTCA 422

RESULT 41
AF140160 531 bp DNA PLN 12-FEB-2001
LOCUS Filobasidiella neoformans isolate CN196.98 diphenol oxidase gene,
DEFINITION partial cds.
ACCESSION AF140160
VERSION AF140160.1 GI:5733998
KEYWORDS
SOURCE
ORGANISM Filobasidiella neoformans.
Filobasidiella neoformans
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
Tremellales; Tremellaceae; Filobasidiella.
1 (bases 1 to 531)
Xu,J., Vilgalys,R. and Mitchell,T.G.
Multiple gene genealogies reveal recent dispersion and
hybridization in the human pathogenic fungus Cryptococcus
neoformans
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE 20504730
PUBMED 11050543
REFERENCE 2 (bases 1 to 531)
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
University, Research Dr., Durham, NC 27710, USA
FEATURES
Source Location/Qualifiers
1..531
/organism="Filobasidiella neoformans"
/isolate="CN196.98"
/db_xref="taxon:5207"
join(<95..154,205..320,375..458,511..>531)
/product="diphenol oxidase"
join(<95..154,205..320,375..458,511..>531)
/note="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD49820.1"
/db_xref="GI:5734032"
/translation="MHGLROLGTAFMGVPGITQCPPIPGSSFTYQFTVSHSGTFMW
HSHYSNMAIGIMGPLITHSPNEPLQGRYDDEDRIVFTIDWHDNSEVY"
BASE COUNT 144 a 122 c 118 g 147 t
ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 531;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcgtcctcgta 18
|||||
Db 436 TTCGATCCTCGTCA 423

```



```

RESULT 42
AF140158/c 532 bp DNA PLN 12-FEB-2001
LOCUS AF140158
DEFINITION Filobasidiella neoformans isolate M0053 diphenol oxidase gene,
partial cds.
ACCESSION AF140158
VERSION AF140158.1 GI:5733996
KEYWORDS
SOURCE
ORGANISM Filobasidiella neoformans.
            Filobasidiella neoformans
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
            Tremellales; Tremellaceae; Filobasidiella.
REFERENCE
AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.
TITLE Multiple gene genealogies reveal recent dispersion and
            hybridization in the human pathogenic fungus Cryptococcus
            neoformans
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE 20504730
PUBMED 11050543
REFERENCE
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
            University, Research Dr., Durham, NC 27710, USA
FEATURES
source
1..532
/organism="Filobasidiella neoformans"
/isolate="M0053"
/db_xref="taxon:5207"
join(<96..155,206..321,376..459,512..>532)
/product="diphenol oxidase"
join(<96..155,206..321,376..459,512..>532)
/note="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD49818.1"
/db_xref="GI:5734030"
/translation="WHGLROLGTAFMDCVPGITOCPIPPGSSFTYQFTVCHOSGTFMW
            HSHVSNMADGIMGPLIHSNPELQGRDYEDRIVITDWHDNSEVY"
BASE COUNT 147 a 122 c 119 g 144 t
ORIGIN
Query Match 56.0%; Score 14; DB 13; Length 532;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ttgcagctcgtca 18
|||||
DB 437 ttgcagctcgtca 424

RESULT 43
AF140159/c 532 bp DNA PLN 12-FEB-2001
LOCUS AF140159
DEFINITION Filobasidiella neoformans isolate CN110.97 diphenol oxidase gene,
partial cds.
ACCESSION AF140159
VERSION AF140159.1 GI:5733997
KEYWORDS
SOURCE
ORGANISM Filobasidiella neoformans.
            Filobasidiella neoformans
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
            Tremellales; Tremellaceae; Filobasidiella.
REFERENCE
AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.
TITLE Multiple gene genealogies reveal recent dispersion and
            hybridization in the human pathogenic fungus Cryptococcus
            neoformans
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE 20504730
PUBMED 11050543
REFERENCE
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
            University, Research Dr., Durham, NC 27710, USA
FEATURES
source
1..532
/organism="Filobasidiella neoformans"
/isolate="CN110.97"
/db_xref="taxon:5207"
join(<96..155,206..321,376..459,512..>532)
/product="diphenol oxidase"
join(<96..155,206..321,376..459,512..>532)
/note="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD49821.1"
/db_xref="GI:5734033"
/translation="WHGLROLGTAFMDCVPGITOCPIPPGSSFTYQFTVCHOSGTFMW
            HSHVSNMADGIMGPLIHSNPELQGRDYEDRIVITDWHDNSEVY"
BASE COUNT 147 a 122 c 119 g 144 t
ORIGIN
Query Match 56.0%; Score 14; DB 13; Length 532;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ttgcagctcgtca 18
|||||
DB 437 ttgcagctcgtca 424

```

```

PUBMED 11050543
REFERENCE
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
            University, Research Dr., Durham, NC 27710, USA
FEATURES
source
1..532
/organism="Filobasidiella neoformans"
/isolate="CN110.97"
/db_xref="taxon:5207"
join(<96..155,206..321,376..459,512..>532)
/product="diphenol oxidase"
join(<96..155,206..321,376..459,512..>532)
/note="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD49819.1"
/db_xref="GI:5734031"
/translation="WHGLROLGTAFMDCVPGITOCPIPPGSSFTYQFTVCHOSGTFMW
            HSHVSNMADGIMGPLIHSNPELQGRDYEDRIVITDWHDNSEVY"
BASE COUNT 145 a 122 c 118 g 147 t
ORIGIN
Query Match 56.0%; Score 14; DB 13; Length 532;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ttgcagctcgtca 18
|||||
DB 437 ttgcagctcgtca 424

RESULT 44
AF140161/c 532 bp DNA PLN 12-FEB-2001
LOCUS AF140161
DEFINITION Filobasidiella neoformans isolate KWS diphenol oxidase gene,
partial cds.
ACCESSION AF140161
VERSION AF140161.1 GI:5733999
KEYWORDS
SOURCE
ORGANISM Filobasidiella neoformans.
            Filobasidiella neoformans
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
            Tremellales; Tremellaceae; Filobasidiella.
REFERENCE
AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.
TITLE Multiple gene genealogies reveal recent dispersion and
            hybridization in the human pathogenic fungus Cryptococcus
            neoformans
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE 20504730
PUBMED 11050543
REFERENCE
AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
            University, Research Dr., Durham, NC 27710, USA
FEATURES
source
1..532
/organism="Filobasidiella neoformans"
/isolate="KWS"
/db_xref="taxon:5207"
join(<96..155,206..321,376..459,512..>532)
/product="diphenol oxidase"
join(<96..155,206..321,376..459,512..>532)
/note="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD49821.1"
/db_xref="GI:5734033"
/translation="WHGLROLGTAFMDCVPGITOCPIPPGSSFTYQFTVCHOSGTFMW
            HSHVSNMADGIMGPLIHSNPELQGRDYEDRIVITDWHDNSEVY"
BASE COUNT 145 a 122 c 118 g 147 t
ORIGIN
Query Match 56.0%; Score 14; DB 13; Length 532;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ttgcagctcgtca 18
|||||
DB 437 ttgcagctcgtca 424

```

BASE COUNT 146 a 126 c 116 g 144 t  
 ORIGIN HSHYTNMSDSIMGPIIVHSPNEPLQGRGDXDEDRIVFTTQVHDSNVV"

Query Match 56.0%; Score 14; DB 13; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgcattcctgca 18  
 ||||||||||||  
 Db 437 TTGCATCTCTGCA 424

RESULT 45  
 AFI40162 532 bp DNA PLN 12-FEB-2001  
 LOCUS AFI40162/c  
 DEFINITION Filobasidiella neoformans isolate CN-A diphenol oxidase gene,  
 partial cds.

ACCESSION AFI40162  
 VERSION AFI40162.1 GI:5734000

KEYWORDS  
 SOURCE Filobasidiella neoformans.  
 ORGANISM Filobasidiella neoformans  
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;  
 Tremellales; Tremellaceae; Filobasidiella.

REFERENCE  
 AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.  
 TITLE Multiple gene genealogies reveal recent dispersion and  
 hybridization in the human pathogenic fungus *Cryptococcus*  
 neoformans  
 JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)

MEDLINE  
 20504730  
 PUBMED 11050543  
 REFERENCE 2 (bases 1 to 532)

AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.  
 TITLE Direct Submission

JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke  
 University, Research Dr., Durham, NC 27710, USA

FEATURES  
 Location/Qualifiers

1..532

/organism="Filobasidiella neoformans"

/isolate="CN-A"

/db\_xref="taxon:5207"

join(<96..155,206..321,376..459,512..>532)

/product="diphenol oxidase"

join(<96..155,206..321,376..459,512..>532)

/note="laccase precursor"

/codon\_start=1

/product="diphenol oxidase"

/protein\_id="AAD49822.1"

/db\_xref="GI:5734034"

/translation="WHGLRGLGTAFMDCVPGTGCPIPPGSSFTYQFTVSHOSGTFWV  
 HSHYSNMAAGIMGPIIVHSPNEPLQGRGDXDEDRIVFTTQVHDSNVV"

BASE COUNT 144 a 122 c 118 g 148 t  
 ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgcattcctgca 18  
 ||||||||||||  
 Db 437 TTGCATCTCTGCA 424

Search completed: October 9, 2001, 15:47:46  
 Job time: 14852 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:52:46 ; Search time 469.84 Seconds  
(without alignments)  
33.410 Million cell updates/sec

Title: US-09-396-196f-8

Perfect score: 25  
Sequence: 1 gcatttcgacccctcgtcagtgacg 25

Scoring table: OLIGO-NUC  
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 24050

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N\_Geneseq\_0601:\*

|     |  |
|-----|--|
| 1:  | /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT:* |
| 2:  | /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT:* |
| 3:  | /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT:* |
| 4:  | /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT:* |
| 5:  | /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT:* |
| 6:  | /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT:* |
| 7:  | /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT:* |
| 8:  | /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT:* |
| 9:  | /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT:* |
| 10: | /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT:* |
| 11: | /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT:* |
| 12: | /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT:* |
| 13: | /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT:* |
| 14: | /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT:* |
| 15: | /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT:* |
| 16: | /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT:* |
| 17: | /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT:* |
| 18: | /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:* |
| 19: | /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:* |
| 20: | /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT:* |
| 21: | /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:* |
| 22: | /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT:* |

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 25    | 100.0       | 1041   | 20    | AAAX01303   |
| 2          | 25    | 100.0       | 1084   | 10    | AAAG1329    |
| 3          | 25    | 100.0       | 1121   | 7     | AAAG6496    |
| 4          | 25    | 100.0       | 5872   | 15    | AAAG62386   |
| 5          | 15    | 60.0        | 34     | 16    | AAAG75832   |
| 6          | 15    | 60.0        | 586    | 21    | AAAG75563   |
| 7          | 14    | 56.0        | 66     | 20    | AAAG19643   |
| 8          | 14    | 56.0        | 974    | 21    | AAAG35170   |
| 9          | 14    | 56.0        | 974    | 21    | AAAG48947   |
| 10         | 14    | 56.0        | 1160   | 21    | AAAG29322   |
| 11         | 14    | 56.0        | 1425   | 22    | AAAG71824   |

|   |    |    |      |         |    |            |
|---|----|----|------|---------|----|------------|
| C | 12 | 14 | 56.0 | 1758    | 19 | AAAG4597   |
| C | 13 | 14 | 56.0 | 2104    | 17 | AAAG29819  |
| C | 14 | 14 | 56.0 | 3720    | 21 | AAAG36323  |
| C | 15 | 14 | 56.0 | 273254  | 21 | AAAG1914   |
| C | 16 | 13 | 52.0 | 351     | 21 | AAAG8252   |
| C | 17 | 13 | 52.0 | 353     | 21 | AAAG5641   |
| C | 18 | 13 | 52.0 | 375     | 21 | AAAG1754   |
| C | 19 | 13 | 52.0 | 470     | 21 | AAAG75310  |
| C | 20 | 13 | 52.0 | 591     | 21 | AAAG54077  |
| C | 21 | 13 | 52.0 | 606     | 21 | AAAG35502  |
| C | 22 | 13 | 52.0 | 623     | 21 | AAAG75197  |
| C | 23 | 13 | 52.0 | 996     | 21 | AAAG47653  |
| C | 24 | 13 | 52.0 | 1114    | 21 | AAAG41953  |
| C | 25 | 13 | 52.0 | 1284    | 19 | AAAG8707   |
| C | 26 | 13 | 52.0 | 1284    | 19 | AAAG2945   |
| C | 27 | 13 | 52.0 | 1359    | 21 | AAAG36484  |
| C | 28 | 13 | 52.0 | 1497    | 21 | AAAG1323   |
| C | 29 | 13 | 52.0 | 1566    | 22 | AAAG70984  |
| C | 30 | 13 | 52.0 | 1782    | 21 | AAAG38498  |
| C | 31 | 13 | 52.0 | 1783    | 21 | AAAG51337  |
| C | 32 | 13 | 52.0 | 1812    | 21 | AAAG53332  |
| C | 33 | 13 | 52.0 | 3172    | 21 | AAAG63854  |
| C | 34 | 13 | 52.0 | 3926    | 21 | AAAG63853  |
| C | 35 | 13 | 52.0 | 4145    | 21 | AAAG237098 |
| C | 36 | 13 | 52.0 | 5520    | 19 | AAAG03311  |
| C | 37 | 13 | 52.0 | 130480  | 22 | AAAG25833  |
| C | 38 | 13 | 52.0 | 1038602 | 20 | AAAG201425 |
| C | 39 | 12 | 48.0 | 19      | 14 | AAAG51957  |
| C | 40 | 12 | 48.0 | 20      | 21 | AAAG6407   |
| C | 41 | 12 | 48.0 | 50      | 14 | AAAG50268  |
| C | 42 | 12 | 48.0 | 61      | 16 | AAAG00337  |
| C | 43 | 12 | 48.0 | 61      | 22 | AAAG70923  |
| C | 44 | 12 | 48.0 | 110     | 19 | AAAG12922  |
| C | 45 | 12 | 48.0 | 110     | 19 | AAAG11656  |

#### ALIGNMENTS

|          |  |               |      |      |     |  |
|----------|--|---------------|------|------|-----|--|
| RESULT 1 |  |               |      |      |     |  |
| ID       | AAAX01303  | standard;     | DNA; | 1041 | BP. |  |
| XX       | AAAX01303;   |               |      |      |     |  |
| AC       | AAAX01303;   |               |      |      |     |  |
| DT       | 12-APR-1999  | (first entry) |      |      |     |  |
| XX       |  |               |      |      |     |  |
| DE       | E. coli biotin synthetase (BiotB) coding sequence.               |               |      |      |     |  |
| KW       | DAP aminotransferase; diaminopelargonic acid; transgenic plant;  |               |      |      |     |  |
| KM       | biotin synthase; biotin production; vitamin H; BiotB; ss.        |               |      |      |     |  |
| XX       |  |               |      |      |     |  |
| OS       | Escherichia coli.  |               |      |      |     |  |
| XX       |  |               |      |      |     |  |
| PN       | US5869719-A.   |               |      |      |     |  |
| XX       |  |               |      |      |     |  |
| PD       | 09-FEB-1999.   |               |      |      |     |  |
| XX       |  |               |      |      |     |  |
| PF       | 30-APR-1997;   | 97US-0846338. |      |      |     |  |
| XX       |  |               |      |      |     |  |
| PR       | 30-APR-1997;   | 97US-0846338. |      |      |     |  |
| PR       | 08-MAR-1995;   | 95US-0401068. |      |      |     |  |
| XX       |  |               |      |      |     |  |
| PA       | (NOVS ) NOVARTIS FINANCE CORP.                                   |               |      |      |     |  |
| XX       |  |               |      |      |     |  |
| PI       | Patton DA;   |               |      |      |     |  |
| XX       |  |               |      |      |     |  |
| PI       | WPI; 1999-152902/13.   |               |      |      |     |  |
| XX       |  |               |      |      |     |  |
| DR       | P-PSDB; AAW73906.  |               |      |      |     |  |
| XX       |  |               |      |      |     |  |
| PT       | Transgenic plants with high biotin levels - transformed with DNA |               |      |      |     |  |
| PT       | encoding di- amino- pelargonic acid amino- transferase or biotin |               |      |      |     |  |
| PT       | synthase   |               |      |      |     |  |

Mouse uncoupling P  
Serine hydroxymeth  
Mechanical stress  
Chlamydia pneumoni  
Human secreted pro  
Euclalyptus grandis  
Human secreted pro  
Human ORF ORF865  
Arabidopsis thalia  
Arabidopsis thalia  
Human ORF ORF752  
Arabidopsis thalia  
Arabidopsis thalia  
DNA encoding a S.  
Streptococcus pneu  
Arabidopsis thalia  
Arabidopsis thalia  
C. glutamicum SRT  
Arabidopsis thalia  
Arabidopsis thalia  
Neisseria gonorrhe  
Nucleotide sequenc  
DNA encoding de no  
Thermococcus 9N-2  
R. marinus bacteri  
Complete genome se  
BCL-2 mRNA ribozym  
Dog genomic marker  
HIV env INS mutage  
Family 2 bFGF DNA  
bFGF DNA ligand #5  
Human biallelic po  
Human biallelic po

XX Example 2; Column 37-40; 34pp; English.

PS This sequence encodes the E. coli biotin synthetase (BioB). The gene can  
 CC be used in the transgenic plant of the invention. The transgenic plant,  
 CC plant cell or plant tissue is transformed with a chimeric gene encoding  
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and  
 CC produces more biotin than a non-transgenic plant, cell or tissue. The  
 CC plant is used as an improved dietary source of biotin (vitamin H) for  
 CC humans or animals.

XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;  
 Best Local Similarity 100.0%; Pred. No. 8.7e-06;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatcgcattcctcgtcagtgacg 25  
 |||||||||||||||||||||||||  
 Db 99 gcatcgcattcctcgtcagtgacg 123

# RESULT 2

AA062386  
 ID AA062386 standard; DNA: 1084 BP.

AC AA062386;

DT 15-FEB-1990 (first entry)

DE E.coli Bio B gene.

KW E.coli; Bio B gene; biotin.

XX Escherichia coli.

OS Key Location/Qualifiers

FT CDS 24..1064

XX /\*tag=a

PN GR2216530-A.

XX 11-OCT-1989.

PF 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

XX Pearson BM, McKee RA;

PI WPI; 1989-295085/41. P-PSDB P91392

DR Plasmid contg. gene(s) for expression of biotin synthetase enzymes

PT - derived from E.coli and capable of replication and expression in other

XX microorganisms, esp. yeast.

XX Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae

CC are plasmids pMA91, pMA36c, pKY49 and pCK495, and plasmid pCK965 for

CC Lactobacillus. Insertion of bio B improves biotin yields in

CC microorganisms which export biotin, or enables growth in media contg.

CC little or no biotin of organisms unable to synthesise biotin for their

CC own use.

XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;  
 Best Local Similarity 100.0%; Pred. No. 8.7e-06;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatcgcattcctcgtcagtgacg 25  
 |||||||||||||||||||||||||  
 Db 122 gcatcgcattcctcgtcagtgacg 146

# RESULT 3

AA060496  
 ID AA060496 standard; DNA: 1121 BP.

AC AA060496;

DT 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetic enzyme; E.coli; desbiobiotin; ds.

XX Key Location/Qualifiers

FT CDS 42..1082

XX /\*tag=a

PN JP61149091-A.

XX 07-JUL-1986.

PF 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

XX (NIPS) NIPPON SODA KK.

XX WPI: 1986-216622/33.

DR P-PSDB; AAP60536.

XX Double stranded DNA encoding biotin synthesising enzyme -

PT comprises transformed mutant E.coli strain contg. cyclic doubled

XX stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

PS Disclosure: Page 534; 23pp; Japanese.

CC The sequence may be expressed by a transformed E.coli host, cultured

XX in a medium containing desbiobiotin.

XX Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

OY 1 gcatcgcattcctcgtcagtgacg 25  
 |||||||||||||||||||||||||  
 Db 140 gcatcgcattcctcgtcagtgacg 164

# RESULT 4

AA062386  
 ID AA062386 standard; DNA: 5872 BP.

AC AA062386;

DT 16-NOV-1994 (first entry)

DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

XX Biotin; expression; enterobacteria; vitamin H; synthesis;

XX plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;

XX promoter; pTac; biotin synthase; KAPA synthase;

XX 8-amino-7-oxononanoate synthase; pimeoyl-CoA; DTB synthase;

XX



```

XX PN EP628572-A.
XX PD 14-DEC-1994.
XX PF 27-MAY-1994; 94EP-0108256.
XX PR 28-MAY-1993; 93JP-0126709.
XX PR 02-MAR-1994; 94JP-0032201.
XX PA (ARIMV) ARIMA T.
XX PA (EISA) EISAI CO LTD.
XX PI Aoyama M, Arima T, Hosoda T, Iwasaki Y, Ohara T;
XX PI Sawada T, Tohmatsu J;
XX DR WPI: 1995-015655/03.
XX PS
XX PT New non-A non-B hepatitis virus sub-type - used to develop prods.
XX PT for detection, diagnosis, prevention and treatment of non-A non-B
XX PT hepatitis.
XX PS Example 2; Page 49; 59pp; English.
XX CC This primer is based on nucleotides 778-800 of AA075818 (a part of
XX CC the Non-A Non-B hepatitis virus genome encoding the non-structural
XX CC protein). It is used in conjunction with AA075833 to amplify nucleotides
XX CC 800-1326 of AA075818. The nucleotide sequences (see also AA075817-19)
XX CC were isolated from the plasma of donors in Japan with high s-gpt levels,
XX CC and were found to be different from previously reported NANB hepatitis
XX CC viruses. The DNA can be used as a reagent for detecting the NANB
XX CC hepatitis viral gene. The polypeptides can be used as reagents for
XX CC detecting anti-NANB hepatitis antibodies or as a NANB hepatitis viral
XX CC vaccine.
XX SQ Sequence 34 BP; 6 A; 9 C; 12 G; 7 T; 0 other;

Query Match 60.0%; Score 15; DB 16; Length 34;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gatctcgtcaggtg 22
   |||||
Db 8 gatctcgtcaggtg 22

RESULT 6
AAC75563/c
ID AAC75563 standard; cDNA; 586 BP.
XX
AC AAC75563;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1118 polynucleotide sequence SEQ ID NO:2235.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerable; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineoplastic;
KW antiviral; antibacterial; antifungal; antineumatic; antihypertoid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antineoplastic disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.

```

```

XX PN WO200058473-A2.
XX PD 05-OCT-2000.
XX PF 31-MAR-2000; 2000MO-US08621.
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PA (CURA-) CURAGEN CORP.
XX PI Shinkets RA, Leach M;
XX PI WPI: 2000-602362/57.
XX DR P-PSDB; AAB41354.
XX PS
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -
XX PS Claim 5; Page 1649; 5507pp; English.
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnerable;
XX CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
XX CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;
XX CC antineoplastic; antibacterial; antiviral; antifungal; antineumatic;
XX CC antihypertoid; and antianaemic. The sequences can be used for determining
XX CC the presence of or predisposition to, or preventing or treating
XX CC pathological conditions associated with an ORFX-associated disorder. The
XX CC nucleic acids can be used to express ORFX proteins in gene therapy.
XX CC vectors. The proteins and nucleic acids may be used to treat cancers,
XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,
XX CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
XX CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX SQ Sequence 586 BP; 122 A; 185 C; 156 G; 123 T; 0 other;

Query Match 60.0%; Score 15; DB 21; Length 586;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atctcgtcaggttc 23
   |||||
Db 304 ATCCTCGTCAGGTGC 290

RESULT 7
AAZ19643
ID AAZ19643 standard; RNA; 66 BP.
XX
AC AAZ19643;
XX
DT 08-NOV-1999 (first entry)
XX
DE Complement system protein C1q RNA binding ligand 29.
XX
KW Complement system protein C1q RNA binding ligand 29.
KW neuroprotective; anti-asthmatic; cerebroprotective; immunosuppressive;
KW diagnostic; prevention; treatment; complement protein-related disease;
KW Alzheimer's disease; renal disease; transplant rejection; asthma; ss.

```

XX OS Synthetic.  
XX PN WO941271-A1.  
XX PD 19-AUG-1999.  
XX PF 05-FEB-1999; 99WO-US02597.  
XX PR 29-SEP-1998; 98US-0163025.  
XX PR 12-FEB-1998; 98US-0023228.  
XX PA (NEXS-) NEXSTAR PHARM INC.  
XX PI Biesecker G, Gold L;  
XX DR WPI; 1999-527357/44.  
XX PT New Nucleic Acid Ligand to complement protein C5, useful in the  
PT treatment of Alzheimer's disease, renal diseases, transplant  
PT rejection, stroke and asthma  
XX PS Example 6; Page 45; 120pp; English.  
XX CC This invention describes novel purified and isolated non-naturally  
CC occurring RNA ligands (I) to complement system proteins C5, C1q and C3.  
CC The products of the invention have neurotropic, neuroprotective,  
CC anti-asthmatic, cerebroprotective and immunosuppressive activity. The  
CC new ligands are useful as diagnostic agents, and pharmaceutical agents  
CC for prevention and treatment of complement protein-related diseases,  
CC including Alzheimer's disease, renal diseases, transplant rejection,  
CC stroke and asthma.  
XX SQ Sequence 66 BP; 11 A; 19 C; 23 G; 13 U; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 66;  
Best Local Similarity 71.4%; Pred. No. 24;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcagctcgtcag 19  
:||||:||||:||||  
Db 40 ucgagccugcag 53

RESULT 8  
AAC35170  
ID AAC35170 standard; DNA; 974 BP.  
XX AC AAC35170;  
XX AC AAC35170;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9244.  
XX KM Hybridisation assay; genetic mapping; gene expression control;  
KM protein identification; signal transduction pathway;  
KM metabolic pathway; promoter; termination sequence; ss.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 25-MAR-1999; 99US-0126264.  
XX PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139819.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.

```
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150560.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 56.0%; Score 14; DB 21; Length 974;
Best Local Similarity 100.0%; Pred No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 ttctgacccgcgc 17
|||||
Db 593 ttctgacccgcgc 606
RESULT 9
AAC48947
ID AAC48947 standard; DNA; 974 BP.
XX
AC AAC48947;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59360.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
```



PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135253.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137528.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 18-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140655.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142350.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144337.  
PR 19-JUL-1999; 99US-0144337.  
PR 19-JUL-1999; 99US-0144337.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-01452913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0148368.  
PR 17-AUG-1999; 99US-0148175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160776.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 56.0%; Score 14; DB 21; Length 974;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 ttgcagtcctgc 17  
 |||||  
 Db 593 ttgcagtcctgc 606

RESULT 10  
 ID AAA29322 standard; cDNA; 1160 BP.  
 XX  
 AC AAA29322;  
 XX  
 DT 26-SEP-2000 (first entry)  
 XX  
 DE Glycine max isoflavone O-methyltransferase clone srl.pk0015.b4.  
 XX  
 KW Isoflavone O-methyltransferase; soybean; degradation; daidzein;  
 KW Flavanoid; flower colour; growth; pollination; irradiation; ss.  
 XX  
 OS Glycine max.  
 XX  
 XX Key Location/Qualifiers  
 FH 11.1087  
 FT CDS /\*tag= a  
 FT /product= Isoflavone-O-methyltransferase  
 XX  
 PN WO200037656-A2.  
 XX  
 PD 29-JUN-2000.  
 PD  
 PF 20-DEC-1999; 99WO-US30338.  
 PF  
 PR 21-DEC-1998; 98US-0113190.  
 PR  
 XX (DUPO ) DU PONT DE MEMOURS & CO E I.  
 PA  
 PI Famodu OO, McGonigle B, Odell JT, Fader GM, Falco SC;  
 DR  
 DR MPI: 2000-442680/38.  
 DR P-PSDB; AAY96584.  
 XX  
 PT New polynucleotide encoding flavonoid biosynthetic enzymes for creating  
 PT transgenic plants and for immunological screening of cDNA libraries  
 XX  
 PS Claim 2: Page 32; 39pp; English.  
 CC AAA29321-25 encode isoflavone O-methyltransferases isolated from various  
 CC soybean (Glycine max) tissues. cDNA libraries seq, srl, srlic and srlic

CC were prepared from soybean embryo (19 days after flowering), root, 8-day  
 CC old root and seed (25 days after flowering).  
 CC Isoflavone O-methyltransferase catalyses the first step in degradation  
 CC of daidzein. Suppression of this enzyme will yield higher concentrations  
 CC of this beneficial isoflavone in, e.g. soybean seed. Flavonoids serve as  
 CC co-pigments in flower colour, stimulate pollen tube growth, attract  
 CC pollinators, act as feeding deterrents and protect against UV  
 CC irradiation in fruits and seeds. The cDNA and proteins can be used to  
 CC isolate homologues, for immunological screening and for positive  
 CC selection methods.

Sequence 1160 BP; 361 A; 197 C; 254 G; 348 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1160;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ctctcagtgtagc 25  
 |||||  
 Db 222 ctctcagtgtagc 235

RESULT 11  
 ID AAF71824 standard; DNA; 1425 BP.  
 XX  
 AC AAF71824;  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:143.  
 XX  
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
 KW fine chemical production; microorganism; organic acid; nucleoside;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
 KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme; ds.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200100843-A2.  
 XX  
 PD 04-JAN-2001.  
 PD  
 PF 23-JUN-2000; 2000WO-1B00923.  
 PF  
 PR 25-JUN-1999; 99US-0141031.  
 PR 01-JUL-1999; 99DE-1030476.  
 PR 02-JUL-1999; 99US-0142101.  
 PR 08-JUL-1999; 99DE-1031415.  
 PR 08-JUL-1999; 99DE-1031418.  
 PR 08-JUL-1999; 99DE-1031419.  
 PR 08-JUL-1999; 99DE-1031420.  
 PR 08-JUL-1999; 99DE-1031424.  
 PR 08-JUL-1999; 99DE-1031428.  
 PR 08-JUL-1999; 99DE-1031434.  
 PR 08-JUL-1999; 99DE-1031435.  
 PR 08-JUL-1999; 99DE-1031443.  
 PR 08-JUL-1999; 99DE-1031453.  
 PR 08-JUL-1999; 99DE-1031457.  
 PR 08-JUL-1999; 99DE-1031465.  
 PR 08-JUL-1999; 99DE-1031478.  
 PR 08-JUL-1999; 99DE-1031510.  
 PR 08-JUL-1999; 99DE-1031541.  
 PR 08-JUL-1999; 99DE-1031573.  
 PR 08-JUL-1999; 99DE-1031592.  
 PR 08-JUL-1999; 99DE-1031632.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 08-JUL-1999; 99DE-1031636.  
 PR 09-JUL-1999; 99DE-1032125.  
 PR 09-JUL-1999; 99DE-1032126.  
 PR 09-JUL-1999; 99DE-1032130.



DE Serine hydroxymethyl transferase gene.  
XX  
KW serine hydroxymethyl transferase; Coryneform bacteria; production; ds.  
XX  
OS Brevibacterium flavum.  
XX  
FH Key Location/Qualifiers  
FT CDS 556..1860  
FT /\*tag= a  
FT /product= serine\_hydroxymethyl\_transferase  
XX  
PN JP08107788-A.  
XX  
PD 30-APR-1996.  
XX  
PF 11-OCT-1994; 94JP-0245435.  
XX  
PR 11-OCT-1994; 94JP-0245435.  
XX  
PA (MITU ) MITSUBISHI CHEM CORP.  
XX  
DR WPI: 1996-262598/27.  
DR P-PSDB: AAR97745.  
XX  
PT Serine hydroxymethyl transferase gene from Brevibacterium flavum  
PT MU-233 - useful for recombinant prodn. of the enzyme in transformed  
PT coryneform bacteria  
XX  
PS Claim 1; Page 5-7; 7pp; Japanese.  
XX  
CC The present sequence encodes a serine hydroxymethyl transferase  
CC isolated from Brevibacterium flavum. The DNA can be used to  
CC transform Coryneform bacterium to produce the enzyme in large  
CC quantities.  
XX  
SQ Sequence 2104 BP; 405 A; 501 C; 610 G; 588 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 2104;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttgcgacgcgcgc 17  
|||  
DB 1635 ttgcgacgcgcgc 1648

RESULT 14  
AAZ36323/c  
ID AAZ36323 standard; cDNA; 3720 BP.  
XX  
AC AAZ36323;  
XX  
DT 22-FEB-2000 (first entry)  
XX  
DE Mechanical stress induced cDNA encoding protein 405.  
XX  
KW Mechanical stress; gene therapy; protein 405; osteoporosis; bone density;  
KW bone development; ss.  
XX  
OS Rattus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 46..3177  
FT /\*tag= a  
FT /product= "protein 405"  
XX  
PN W09960164-A1.  
XX  
PD 25-NOV-1999.  
XX  
PF 14-MAY-1999; 99WO-US11066.  
XX

PR 15-MAY-1998; 98US-0085673.  
XX  
PA (QUAR-) QUARK BIORECH INC.  
XX  
PI Einat P, Mor O, Skallier R, Feinstein E, Faerman A;  
XX  
DR WPI: 2000-053304/04.  
DR P-PSDB: AAY53669.  
XX  
PT Identification of stress induced genes for determining risk and  
PT preventing, treating or controlling osteoporosis  
XX  
PS Claim 25; Fig 9A-M; 308pp; English.  
XX  
CC The present sequence encodes protein 405, which was identified  
CC using the method of the invention after subjecting rat osteoblasts to  
CC mechanical stress. Expression of the 405 gene was found to be  
CC downregulated in response to mechanical stimulation and stress.  
CC The specification describes a method for the identification of genes  
CC responsive to a specific mechanical stress. The method comprises applying  
CC the mechanical stress to an organism (tissue or cells comprising bone  
CC cells), isolating the specific cellular fractions and extracting mRNA  
CC from them, and differentially analysing the mRNA in comparison with  
CC control samples. The method is used to identify genes whose expression  
CC is responsive to a specific stress. The identified genes are employed in  
CC determining risk associated with a physiological or disease state. The  
CC risk determination methods are used for testing a medicament for gene  
CC therapy. These medicaments, or genes identified by the method of the  
CC invention, are used for treating, preventing or controlling a  
CC physiological or disease state (especially osteoporosis or bone density  
CC or other factors causing or contributing to osteoporosis or its symptoms  
CC or other conditions involved in mechanical stress or its lack. The  
CC methods can also be used for advancing research or studies in bone  
CC development.  
XX  
SQ Sequence 3720 BP; 946 A; 991 C; 955 G; 828 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 3720;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gatccgcgcgcgcgc 21  
|||  
DB 574 GATCCTCGTCAGGT 561

RESULT 15  
AAC81914  
ID AAC81914 standard; DNA; 273254 BP.  
XX  
AC AAC81914;  
XX  
DT 27-FEB-2001 (first entry)  
XX  
DE Chlamydia pneumoniae genome DNA.  
XX  
KW Genome; diagnosis; vaccine; ds.  
XX  
OS Chlamydia pneumoniae.  
XX  
PN W0200027994-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 12-NOV-1999; 99WO-US26923.  
XX  
PR 12-NOV-1998; 98US-0108279.  
PR 08-APR-1999; 99US-0128606.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Stephens R, Mitchell W, Kalman S, Davis R;

XX WPI: 2000-376516/32.  
 XX  
 PT Isolated nucleic acid for use in diagnostic and analytical methods  
 PT encodes genomic sequence of Chlamydia pneumoniae -  
 XX  
 XX Claim 2: Page 128-320; 320pp; English.  
 PS  
 CC This invention describes a novel nucleic acid (N1) encoding a Chlamydia  
 CC pneumoniae protein (P1), given in the specification. The isolated nucleic  
 CC acid is useful for diagnostic and analytical methods, such as,  
 CC hybridization-based assays or amplification-based assays. The protein may  
 CC be used for diagnostic purposes, for their enzymatic or structural  
 CC activity, or as a vaccine. The invention also describes (1) a probe  
 CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid  
 CC (N2) that hybridizes under stringent conditions to N1; (3) an expression  
 CC cassette comprising N1 under the transcriptional regulation of a  
 CC transcriptional initiation region functional in an expression host, and a  
 CC transcriptional termination region; (4) a cell comprising an expression  
 CC cassette of (3) as part of an extrachromosomal element or integrated into  
 CC the genome of a host cell as a result of induction of the expression  
 CC cassette into the host cell, and the cellular progeny of the host cell;  
 CC (5) a method for producing a P1 comprising growing a cell of (4) where  
 CC the protein is expressed and isolating the protein free of other  
 CC proteins; (6) a purified polypeptide composition comprising at least 50  
 CC weight % of P1; and (7) a monoclonal antibody binding specifically to the  
 CC peptide of (6).  
 CC  
 XX Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other:  
 SQ  
 Query Match 56.0%; Score 14; DB 21; Length 273254;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 gatctctgcaggt 21  
 |||||  
 DB 32246 gatctctgcaggt 32259  
 RESULT 16  
 AAC08252/c  
 ID AAC08252 standard; CDNA; 351 BP.  
 AC AAC08252;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 12327.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP103401-A2.  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GENE-) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI: 2000-500381/45.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX

PS Claim 1: SEQ ID 12327; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 CC  
 XX Sequence 351 BP; 58 A; 105 C; 122 G; 66 T; 0 other:  
 SQ  
 Query Match 52.0%; Score 13; DB 21; Length 351;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 tctctgcaggt 22  
 |||||  
 DB 220 TCCTGTCAGGTG 208  
 RESULT 17  
 AAC56641/c  
 ID AAC56641 standard; DNA; 353 BP.  
 AC AAC56641;  
 XX  
 DT 25-JAN-2001 (first entry)  
 XX  
 DE Euca1yplus grandis transcription factor DNA sequence #512.  
 XX  
 KW Plant; transcription factor; gene expression; euca1yplus; pine; acacia;  
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox;  
 KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;  
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.  
 XX  
 OS Euca1yplus grandis.  
 XX  
 PN W0200053724-A2.  
 PD 14-SEP-2000.  
 XX  
 PF 09-MAR-2000; 2000WO-US06112.  
 XX  
 PR 11-MAR-1999; 99US-026513.  
 XX  
 PR 18-AUG-1999; 99US-0149485.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX  
 PI Wood M, McGrath A, Shenk MA, Glenn M;  
 XX  
 DR WPI: 2000-579369/54.  
 XX  
 PT New isolated polynucleotide encoding a plant transcription factor for  
 PT producing a plant e.g. a woody plant, preferably euca1yplus or pine,  
 PT having modified gene expression or modified activity of a polypeptide  
 PT -  
 XX  
 PS Claim 1: Page 482; 747pp; English.  
 XX  
 CC The present invention relates to novel plant transcription factors from  
 CC Euca1yplus grandis or Pinus radiata. The present sequence is the coding  
 CC sequence for one such transcription factor. The transcription factor may

CC be used to produce a plant having modified gene expression such as a  
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
 CC mahogany species or to modify the activity of a polypeptide in a plant.  
 CC The transcription factors of the present invention are members from the  
 CC following families of regulatory proteins: bZIP, bZIP family of G-box  
 CC binding factors, basic helix-loop-helix zipper,  
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2  
 CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements  
 CC and MYB.  
 CC  
 SO Sequence 353 BP; 117 A; 59 C; 98 G; 78 T; 1 other;

Query Match 52.0%; Score 13; DB 21; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 13 tctctcaggtcagc 25  
 |||  
 Db 326 TCGTCAGGTGCGAG 314

RESULT 18  
 AAC01754/C  
 ID AAC01754 standard; cDNA: 375 BP.  
 XX  
 AAC01754:  
 XX  
 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 1752.

XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.

XX  
 PN EPI033401-A2.  
 PD  
 XX 06-SEP-2000.  
 PD  
 XX 21-FEB-2000; 2000EP-0200610.  
 PF  
 XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.  
 XX  
 PI Dunas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI: 2000-500381/45.  
 DR P-PSDB: AAG01748.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PT  
 PS Claim 1; SEQ ID 1752; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.  
 CC

SO Sequence 375 BP; 66 A; 121 C; 117 G; 62 T; 9 other;

Query Match 52.0%; Score 13; DB 21; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 10 tctctcaggtcagc 22  
 |||  
 Db 182 TCCTCCTCAGGTG 170

RESULT 19  
 AAC75310/C  
 ID AAC75310 standard; cDNA: 470 BP.  
 XX  
 AAC75310:  
 XX  
 08-FEB-2001 (first entry)  
 XX  
 DE Human ORF865 polynucleotide sequence SEQ ID NO:1729.

XX Human: open reading frame; ORF; detection; cytostatic; hepatotropic;  
 KW vulnery; antiparasitic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertoid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.  
 OS  
 XX  
 PN WO200058473-A2.  
 PD  
 XX 05-OCT-2000.  
 PD  
 XX 31-MAR-2000; 2000WO-US08621.  
 PF  
 XX 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.  
 PA  
 XX  
 PI Shinkets RA, Leach M;  
 XX  
 DR WPI: 2000-602362/57.  
 DR P-PSDB: AAB41101.

XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PT  
 PS Claim 5; Page 1368-1369; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antihypertoid; antianemic; The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating

CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
XX Sequence 470 BP; 93 A; 166 C; 131 G; 77 T; 3 other;

Query Match 52.0%; Score 13; DB 21; Length 470;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 tcgattctctgtca 18  
Db 68 TCGATCCTCGTCA 56

## RESULT 20

AAC54077  
ID AAC54077 standard; DNA; 591 BP.  
AC AAC54077;  
XX  
XX 18-OCT-2000 (first entry)  
DT  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 76608.  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KM protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX EP1033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX  
XX 25-FEB-1999; 990S-0121825.  
PR 05-MAR-1999; 990S-0123180.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 23-APR-1999; 990S-0130891.  
PR 28-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 30-APR-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 06-MAY-1999; 990S-0132487.  
PR 07-MAY-1999; 990S-0132863.  
PR 11-MAY-1999; 990S-0134256.  
PR 14-MAY-1999; 990S-0134318.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.

PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142155.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142927.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145091.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.

```
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150864.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160860.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 591;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 attcgatcctcg 15
Db 144 attcgatcctcg 156

RESULT 21
ID AAC35502 standard; DNA; 606 BP.
AC AAC35502;
XX
XX 17-OCT-2000 (first entry)
DT
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10429.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
```



PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143524.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145214.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161932.  
PR 28-OCT-1999; 99US-0161932.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 606;  
 Best Local Similarity 100.0%; Pred. No. 79;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 attcgatccctgcg 15  
 |||||  
 Db 159 attcgatccctgcg 171

RESULT 22  
 AAC75197  
 ID AAC75197 standard; cDNA; 623 BP.  
 AC AAC75197;  
 XX  
 DT 08-FEB-2001 (first entry)

Human ORF752 polynucleotide sequence SEQ ID NO:1503.  
 DE  
 XX Human; open reading frame; ORF; detection; cytosolic; hepatotropic;  
 KW vulnary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antineoplastic; coagulation;  
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.  
 OS  
 XX  
 PN W0200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shimkels RA, leach M;  
 DR WPI: 2000-602362/57.  
 DR P-PSDB: AAB40988.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 5; Page 1245; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORF open reading frames 1 to 3161. The ORF  
 CC sequences have activities such as: cytosolic; hepatotropic; vulnary;  
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antineoplastic; antianemic; antirheumatic; antihypertensive;  
 CC antihypertensive; antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORF-associated disorder. The  
 CC nucleic acids can be used to express ORF proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antineoplastic; coagulation; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX  
 SQ Sequence 623 BP; 119 A; 201 C; 184 G; 119 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 623;  
 Best Local Similarity 100.0%; Pred. No. 79;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ctgcaggtgca 24  
 |||||  
 Db 563 ctgcaggtgca 575

RESULT 23  
 AAC7663/C  
 ID AAC7663 standard; DNA; 996 BP.  
 AC AAC7663;  
 XX  
 DT 18-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 54655.  
 DE  
 XX  
 KW Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.  
 OS  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140053.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142830.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 52.0%; Score 13; DB 21; Length 996;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 4 ttccatcctcct 16  
Db 714 TTTCATCTCTCT 702  
  
RESULT 24  
AAC41953/C  
ID AAC41953 standard; DNA; 1114 BP.  
XX AAC41953;  
XX AC  
XX 17-OCT-2000 (first entry)  
XX DT  
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 33748.  
XX KW Hybridisation assay; genetic mapping; gene expression control;  
XX KW protein identification; signal transduction pathway;  
XX KW metabolic pathway; promoter; termination sequence; ss.  
XX OS Arabidopsis thaliana.  
XX PN EPI033405-A2.  
XX PD 06-SEP-2000.  
XX FE 25-FEB-2000; 2000EP-0301439.  
XX XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130049.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154479.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161932.  
 PR 28-OCT-1999; 99US-0161933.  
 PR 29-OCT-1999; 99US-0162142.

Query Match

52.0%;

Score 13;

DB 21;

Length 1114;

Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 ttgcattcctgt 16  
 |||||  
 Db 674 ttgcattcctgt 662

RESULT 25  
 AAT98707/C  
 ID AAT98707 standard; DNA: 1284 BP.  
 XX  
 AC AAT98707;  
 XX  
 DT 09-NOV-1998 (first entry)  
 XX  
 DE DNA encoding a S. pneumoniae protein of unknown function.  
 XX  
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
 KW immunological response; inoculation; antibody production; inhibitor;  
 KW T cell immune response; antimicrobial compound; bacterial adhesion;  
 KW extracellular matrix protein; protein-mediated cell invasion; wound;  
 KW pathogenesis; ss.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 FH Key location/Qualifiers  
 FT CDS 310..537  
 FT /\*tag= a

W09743303-A1.  
 20-NOV-1997.  
 14-MAY-1997; 97WO-US07950.  
 XX  
 PR 14-MAY-1996; 96US-0017670.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;  
 PI Stodola RK;  
 DR WPI: 1996-008793/01.  
 DR P-PSDB: AAW38663.  
 XX  
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for  
 PT diagnosing anti-microbial agents for treatment of bacterial  
 PT infections  
 PS  
 PS Claim 4: Pages 210-211; 483pp; English.

This sequence encodes a Streptococcus pneumoniae protein of unknown function, and represents a DNA sequence of the invention.  
 The DNA sequences were isolated from Streptococcus pneumoniae strain 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the invention can be used to identify compounds which interact with and inhibit or activate the activity of the proteins. Antagonists can be used to treat diseases caused by S. pneumoniae proteins, through genetic immunisation. They can also be used to induce an immunological response in a mammal by inoculation with the S. pneumoniae proteins or delivery of the encoding nucleic acids in a vector adequate to produce antibody and/or T cell immune responses to protect the animal from disease. The proteins can also be used to identify antimicrobial compounds which are capable of inhibiting their bioactivity. In particular the proteins of the invention can be used to prevent adhesion of bacteria to mammalian extracellular matrix proteins on in-dwelling devices or in wounds, to block protein-mediated mammalian cell invasion, and to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or other surgical techniques.

Sequence 1284 BP; 382 A; 257 C; 320 G; 324 T; 1 other;

Query Match 52.0%; Score 13; DB 19; Length 1284;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tcgattccctgcgta 18  
|||||  
DB 995 TCGATCCTCGTCA 983

RESULT 26  
AAV42945/C  
ID AAV42945 standard; DNA: 1284 BP.

AC AAV42945;

DT 09-NOV-1998 (first entry)

DE Streptococcus pneumoniae polypeptide coding region.

KW Polypeptide; ORF: open reading frame; infection; bacterial;  
streptococcal; bacteremia; diagnosis; prophylaxis; ds.

OS Streptococcus pneumoniae.

EH Key Location/Qualifiers  
FT CDS 714..1049  
/\*tag= a

PN W09823631-A1.

PD 04-JUN-1998.

PE 24-NOV-1997; 97WO-US21976.

PR 27-NOV-1996; 96US-0031879.

PA (SMIK ) SMITHKLINE BEECHAM CORP.  
(SMIK ) SMITHKLINE BEECHAM PLC.

PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;

PI Reid RH, Zarfos PN;

DR WPI: 1998-322654/28.

DR P-PSDB: AAM62663.

PT Streptococcus pneumoniae polynucleotides - useful for developing  
products for diagnosis, prevention and treatment of infections e.g.

PT pneumonia, bacteremia, meningitis or endocarditis

PS Claim 1; Page 35-36; 181pp; English.

CC The sequence is that of a Streptococcal polypeptide coding region.

CC The polypeptide can potentially be used for the diagnosis and  
prevention of bacterial infections, especially sp infection.

CC It may be used for the treatment of diseases such as otitis media,  
conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural  
empyema, endocarditis or infection of the cerebrospinal fluid.

CC Sequence 1284 BP; 382 A; 257 C; 320 G; 324 T; 1 other;

Query Match 52.0%; Score 13; DB 19; Length 1284;  
Best Local Similarity 100.0%; Pred. No. 76;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tcgattccctgcgta 18  
|||||  
DB 995 TCGATCCTCGTCA 983

RESULT 27

ACAC36484  
ID AAC36484 standard; DNA: 1359 BP.  
XX  
AC AAC36484;  
XX  
DT 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 13980.  
XX  
DE Hybridisation assay; genetic mapping; gene expression control;  
KM protein identification; signal transduction pathway;  
KM metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PF 25-FEB-1999; 99US-0121825.  
XX  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.

```
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144351.
PR 19-JUL-1999; 99US-0144352.
PR 19-JUL-1999; 99US-0144353.
PR 19-JUL-1999; 99US-0144354.
PR 19-JUL-1999; 99US-0144355.
PR 19-JUL-1999; 99US-0144356.
PR 20-JUL-1999; 99US-0144357.
PR 20-JUL-1999; 99US-0144652.
PR 20-JUL-1999; 99US-0144654.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146390.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155186.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 1359;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 tcgtcaggtgcag 25
    |||||
Db 24 tcgtcaggtgcag 36

RESULT 28
AAC51323
ID AAC51323 standard; DNM; 1497 BP.
AC AAC51323;
XX
XX
DT 18-OCT-2000 (first entry)
XX
```

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 68089.  
XX  
XX Hybridisation assay: genetic mapping; gene expression control;  
KW Protein identification: signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
PN EPI03405-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0121180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 26-JUN-1999; 99US-0140695.  
PR 29-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145324.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145513.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148345.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.



```

PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155133.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-015865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      52.0%; Score 13; DB 21; Length 1497;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 10 tctctgcagtg 22
    |||||
Db 174 tctctgcagtg 186

```

```

RESULT 29
AAAF0984
ID AAF70984 standard; DNA; 1566 BP.
XX
AC AAF70984;
XX
DT 30-APR-2001 (first entry)
XX
DE C. glutamicum SRT protein nucleotide sequence SPQ ID NO:1.
XX
KW Corynebacterium glutamicum; stress; resistance; tolerance; SRT;
    line chemical production; organic acid; proteinogenic amino acid;
    nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
    nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
    carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW

```

```

KW evolutionary study; environmental hazard; fermentation; ds.
XX
OS Corynebacterium glutamicum.
XX
WO200100804-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00922.
XX
PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030429.
PR 01-JUL-1999; 99US-0142692.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031541.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032914.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99US-0151214.
PR 31-AUG-1999; 99DE-1041382.
XX
PA (BADI ) BASF AG.
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberman G, Lee H;
PI Kim H;
XX WPI: 2001-061972/07.
XX P-PSDB: AAB78871.
XX
PT New isolated Corynebacterium glutamicum nucleic acid encoding a stress,
PT tolerance or resistance protein, for production or modulation of
PT production of fine chemicals, such as, e.g. amino acids, lipids,
PT carbohydrates, or enzymes -
XX
PS Claim 3; Page 111-113; 526pp; English.
XX
CC AAF70984 to AAF71133 encode the Corynebacterium glutamicum stress,
CC resistance, and tolerance (SRT) proteins given in AAB78871 to AAB79020.
CC The C. glutamicum SRT genes (I) can be used in vectors (II) for
CC expression in host cells and production of fine chemicals, such as, an
CC organic acid, a proteinogenic or nonproteinogenic amino acid (preferred),
CC a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a
CC saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic
CC compound, a vitamin, a cofactor, a polyketide, or an enzyme. The fine
CC chemical production can be modulated. The presence of (I) or the SRT
CC proteins (II) encoded by them are used for diagnosing the presence
CC or activity of Corynebacterium diptheriae, (I), (II), (III) and host
CC cells containing them can be used to map the genomes of organisms related
CC to C. glutamicum, to identify and localise C. glutamicum sequences of
CC interest, in evolutionary studies, in determination of SRT protein
CC regions required for function, in modulating the SRT protein activity,
CC and in modulating the activity of an SRT pathway. (II) are used to permit
CC C. glutamicum to survive in an environment that is normally
CC environmentally or chemically hazardous to it. (I) and protein molecules
CC encoded by it increase the survival of C. glutamicum to chemical and
CC environmental hazards and provide a means for continued growth and
CC multiplication in large scale fermentative growth conditions. By
CC increasing the growth rate or maintaining a normal growth rate in poor or
CC toxic conditions, the yield, production and/or efficiency or production
CC of fine chemicals from a culture may be increased.
XX
SQ Sequence 1566 BP; 275 A; 406 C; 403 G; 482 T; 0 other;

```

```

Query Match      52.0%; Score 13; DB 22; Length 1566;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 4 ttctgactcgt 16
    |||||
Db 171 ttctgactcgt 183

```

RESULT 30  
AAC38498  
ID AAC38498 standard; DNA: 1782 BP.  
XX  
AC AAC38498;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 21191.  
XX  
KM Hybridisation assay: genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135525.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144864.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 1782;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctctgcagtg 22  
|||||  
Db 299 tctctgcagtg 311

RESULT 31  
AAC51337  
ID AAC51337 standard; DNA; 1783 BP.  
XX  
AC AAC51337;

XX 18-OCT-2000 (first entry)  
DT  
XX  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 68143.  
DE  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
OS  
PN EP1033405-A2.  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 23-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139457.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.

```

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142380.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144332.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145214.
PR 23-JUL-1999; 99US-0145228.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 1783;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctcgtcagctg 22
Db 299 tctcgtcagctg 311

RESULT 32
AAZ53332/c
ID AAZ53332 standard; DNA; 1812 BP.
XX
AC AAZ53332;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrhoeae ORF 151 partial DNA sequence SPQ ID NO:613.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

```

KW antibacterial; gene therapy; ds.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 PN W09957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHTR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M,  
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M,  
 PI Tettelin H, Venter JC;  
 DR WPI: 2000-062150/05.  
 DR P-PSDB: AAY74570.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX  
 PS Claim 7; Page 430; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941  
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 CC  
 SO Sequence 1812 BP; 458 A; 544 C; 491 G; 319 T; 0 other;  
 XX  
 XX  
 Query Match 52.0%; Score 13; DB 21; Length 1812;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 13 tcgtcaggtcag 25  
 |||||||||  
 Db 1099 TCGTCAGCTGCAG 1087  
 |||||||||  
 RESULT 33  
 ID AAA63854 standard; cDNA; 3172 BP.  
 AC AAA63854;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE Nucleotide sequence of diacylglycerol kinase beta splice variant.  
 XX  
 KW Human: diacylglycerol kinase beta; DAGbeta; diacylglycerol; DAG;  
 KW phosphatidic acid; DAG-dependent protein kinase C activation;  
 KW mood disorder; epilepsy; neurodegenerative disorder; anxiety;  
 KW mood disorder; epilepsy; neurodegenerative disorder; anxiety;

KW schizophrenia; migraine; drug dependence; stroke; Alzheimer's dementia;  
 KW Parkinson's disease; splice variant; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..2322  
 FT /\*tag= a  
 FT /product= "diacylglycerol kinase beta splice  
 variant (hDAGkbeta-SV)"  
 FT  
 FT W0200047723-A2.  
 PN 17-AUG-2000.  
 XX  
 PD 23-DEC-1999; 99WO-GB04421.  
 XX  
 PF 15-FEB-1999; 99GB-0003430.  
 XX  
 PR (GLAX ) GLAXO GROUP LTD.  
 XX  
 PA Caricasole A, Caldara F, Sala CF;  
 XX  
 PI WPI: 2000-506093/45.  
 DR P-PSDB: AAB08200.  
 DR  
 XX  
 PT New human diacylglycerol kinase beta (hDAGkbeta) protein and its  
 PT modulating compounds, useful for treatment of neurodegenerative and  
 PT mood disorders -  
 XX  
 PS Claim 5; Page 32-34; 57pp; English.  
 XX  
 CC The present sequence encodes a splice variant of human diacylglycerol  
 CC kinase beta (DAGkbeta), designated hDAGkbeta-SV. DAG converts  
 CC diacylglycerol (DAG) to phosphatidic acid and attenuates DAG-dependent  
 CC protein kinase C activation. Compounds that modulate the activity of  
 CC DAGkbeta may be administered to a human patient for the treatment or  
 CC prophylaxis of a disorder that is responsive to modulation of DAGK  
 CC activity. The disorder may be a mood disorder, epilepsy, a  
 CC neurodegenerative disorder, anxiety, schizophrenia, migraine, drug  
 CC dependence, stroke, Alzheimer's dementia or Parkinson's disease.  
 CC  
 SO Sequence 3172 BP; 1005 A; 612 C; 685 G; 870 T; 0 other;  
 XX  
 XX  
 Query Match 52.0%; Score 13; DB 21; Length 3172;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 9 atctctgcaggt 21  
 |||||||||  
 Db 1391 atctctgcaggt 1403  
 |||||||||  
 RESULT 34  
 ID AAA63853 standard; cDNA; 3926 BP.  
 AC AAA63853;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE Nucleotide sequence of human diacylglycerol kinase beta (DAGkbeta).  
 XX  
 KW Human: diacylglycerol kinase beta; DAGkbeta; diacylglycerol; DAG;  
 KW phosphatidic acid; DAG-dependent protein kinase C activation;  
 KW mood disorder; epilepsy; neurodegenerative disorder; anxiety;  
 KW schizophrenia; migraine; drug dependence; stroke; Alzheimer's dementia;  
 KW Parkinson's disease; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers

FT CDS 1..2415  
 FT /\*tag= a  
 FT /product= "diacylglycerol kinase beta (DAGKbeta)"  
 XX  
 FN WO200047723-A2.  
 PD 17-AUG-2000.  
 XX  
 XX 23-DEC-1999; 99WO-GB04421.  
 PE  
 XX 15-FEB-1999; 99GB-0003430.  
 PR  
 XX (GLAXO ) GLAXO GROUP LTD.  
 PA  
 XX Caricasole A, Caldara F, Sala CF;  
 PI WPL; 2000-506093/45.  
 DR P-PSDB; AAB08199.  
 XX  
 XX New human diacylglycerol kinase beta (hDAGKbeta) protein and its  
 PT modulating compounds, useful for treatment of neurodegenerative and  
 PT mood disorders -  
 PS Claim 5; Page 29-31; 57pp; English.  
 XX  
 CC The present sequence encodes a human diacylglycerol kinase beta  
 CC (DAGKbeta) polypeptide. DAGK converts diacylglycerol (DAG) to  
 CC phosphatidic acid and attenuates DAG-dependent protein kinase C  
 CC activation. Compounds that modulate the activity of DAGKbeta may  
 CC be administered to a human patient for the treatment of propylaxis  
 CC of a disorder that is responsive to modulation of DAGK activity.  
 CC The disorder may be a mood disorder, epilepsy, a neurodegenerative  
 CC disorder, anxiety, schizophrenia, migraine, drug dependence, stroke,  
 CC Alzheimer's dementia or Parkinson's disease.  
 XX  
 SQ Sequence 3926 BP; 1225 A; 776 C; 820 G; 1105 T; 0 other:  
  
 Query Match 52.0%; Score 13; DB 21; Length 3926;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 9 atcctcgtcaggt 21  
 |||||  
 Db 1391 atcctcgtcaggt 1403  
  
 RESULT 35  
 AA237098  
 ID AA237098 standard; DNA: 4145 BP.  
 XX  
 AC AA237098;  
 XX  
 DT 27-MAR-2000 (first entry)  
 XX  
 DE DNA encoding de novo DNA cytosine methyltransferase DNMT3B1.  
 XX  
 KM De novo DNA cytosine methyltransferase; DNMT3B1; neoplastic disorder;  
 KW carcinoma; sarcoma; leukaemia; DNA methylation; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 XX  
 FH Location/Qualifiers  
 FT CDS 115..2676  
 FT /\*tag= a  
 FT /product= "de novo DNA cytosine methyltransferase  
 DNMT3B1"  
 FT exon 1..108  
 FT /\*tag= b  
 FT /number= 1  
 FT exon 109..256  
 FT /\*tag= c  
 FT /number= 2

FT exon 257..318  
 FT /\*tag= d  
 FT /number= 3  
 FT exon 319..420  
 FT /\*tag= e  
 FT /number= 4  
 FT exon 421..546  
 FT /\*tag= f  
 FT /number= 5  
 FT exon 547..768  
 FT /\*tag= g  
 FT /number= 6  
 FT exon 769..927  
 FT /\*tag= h  
 FT /number= 7  
 FT exon 928..1035  
 FT /\*tag= i  
 FT /number= 8  
 FT exon 1036..1180  
 FT /\*tag= j  
 FT /number= 9  
 FT exon 1181..1240  
 FT /\*tag= k  
 FT /number= 10  
 FT exon 1241..1366  
 FT /\*tag= l  
 FT /number= 11  
 FT exon 1367..1411  
 FT /\*tag= m  
 FT /number= 12  
 FT exon 1412..1491  
 FT /\*tag= n  
 FT /number= 13  
 FT exon 1492..1604  
 FT /\*tag= o  
 FT /number= 14  
 FT exon 1605..1788  
 FT /\*tag= p  
 FT /number= 15  
 FT exon 1789..1873  
 FT /\*tag= q  
 FT /number= 16  
 FT exon 1874..2019  
 FT /\*tag= r  
 FT /number= 17  
 FT exon 2020..2110  
 FT /\*tag= s  
 FT /number= 18  
 FT exon 2111..2259  
 FT /\*tag= t  
 FT /number= 19  
 FT exon 2260..2345  
 FT /\*tag= u  
 FT /number= 20  
 FT exon 2346..2415  
 FT /\*tag= v  
 FT /number= 21  
 FT exon 2416..2534  
 FT /\*tag= w  
 FT /number= 22  
 FT exon 2535..4145  
 FT /\*tag= x  
 FT /number= 23  
  
 WO9967397-A1.  
 PD 29-DEC-1999.  
 XX  
 XX 25-JUN-1999; 99WO-US14373.  
 PF  
 XX 25-JUN-1998; 98US-0090906.  
 PR 24-JUL-1998; 98US-0093993.  
 XX

PA (GEO) GEN HOSPITAL CORP.  
 XX  
 PI Li E, Okano M, Xie S;  
 XX  
 DR WPI: 2000-106298/09.  
 XX  
 P-PSDB: AAV34058.  
 XX  
 PT New mouse and human polypeptides, useful to treat and diagnose  
 PT neoplastic disorders e.g. carcinomas, sarcomas and leukemias  
 XX  
 PS Example 1; Fig 1D; 114pp; English.  
 XX  
 CC The present sequence encodes a human de novo DNA cytosine  
 CC methyltransferase designated DNMT3B1. The DNMT3B gene also produces,  
 CC through alternate splicing, at least two shorter isoforms termed  
 CC DNMT3B2 (comprising amino acids 1-355 and 376-853 of AAV54058, see  
 CC AAV54059, and DNMT3B3 (comprising amino acids 1-355, 376-743 and 807-853  
 CC of AAV54058, see AAV54060) (sequences not given in the specification).  
 CC The polypeptides can be administered therapeutically, especially by  
 CC expressing encoding polynucleotides, to treat diseases associated with  
 CC DNA cytosine methyltransferase, such as neoplastic disorders e.g.  
 CC carcinomas, sarcomas and leukemias. They can be used to diagnose, or  
 CC determine susceptibility to neoplastic disorders, by assaying for  
 CC polypeptide expression levels in mammalian cells/body fluids. They are  
 CC useful to screen for compounds inhibiting/activating the polypeptide. The  
 CC polypeptides can also be used for in vitro de novo methylation of DNA.  
 CC Such in vitro methylation may be used to direct or regulate DNA  
 CC expression in biological systems, e.g. recombinant DNA methylated in  
 CC vitro may be introduced into a cell/organism to increase or decrease  
 CC expression of a desired polypeptide for which the native DNA is  
 CC under-methylated or not methylated. The polypeptides can also be used to  
 CC produce antibodies which are useful to detect and purify the polypeptide  
 CC or therapeutically e.g. to treat neoplastic disorders. The  
 CC polynucleotides are useful to produce probes and primers which are useful  
 CC diagnostically.  
 CC  
 SQ Sequence 4145 BP; 1041 A; 1083 C; 1096 G; 925 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 4145;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tcatccctcgta 18  
 |||||||||||||  
 Db 172 tcatccctcgta 184

RESULT 36  
 AAV03311/C  
 ID AAV03311 standard; DNA: 5520 BP.  
 XX  
 AC AAV03311;  
 XX  
 DT 22-JUN-1998 (first entry)  
 XX  
 DE Thermococcus 9N-2 phosphatase (21ph1) gene.  
 XX  
 KW Alkaline phosphatase; thermostable enzyme; thermophilic bacterium;  
 KW food; detergent; baking; ss.  
 XX  
 OS Thermococcus sp. strain 9N-2.  
 XX  
 PN W09748416-A1.  
 XX  
 PD 24-DEC-1997.  
 XX  
 PF 19-JUN-1997; 97WO-US10784.  
 XX  
 PR 19-JUN-1996; 96US-0033752.  
 XX  
 PA (RECO-) RECOMBINANT BIOCATALYSIS INC.

PI Bylina E, Lee E, Mathur EJ;  
 XX  
 DR WPI: 1998-062851/06.  
 XX  
 PT Thermostable phosphatase(s) - useful in pharmaceutical, food,  
 PT detergent, and baking industries  
 XX  
 PS Claim 12; Page 77-81; 128pp; English.  
 XX  
 CC This polynucleotide encodes a thermostable phosphatase, designated  
 CC 31ph1, of Thermococcus 9N-2. The invention relates to claimed  
 CC polynucleotides (see AAV03301-20) coding for novel thermostable  
 CC phosphatases (see AAV42380-95). Claimed vector and host cells are  
 CC used to produce the enzymes, which can be used in a claimed method  
 CC to hydrolyse phosphate bonds. They can also be used in enzyme  
 CC labeling processes, in certain recombinant DNA techniques, in  
 CC ELISA immunoassays, in enzyme linked gene probes, in research  
 CC applications for removing 5' phosphates in polynucleotides prior to  
 CC end labeling and in the pharmaceutical, food, detergent, and  
 CC baking industries. Polynucleotides can also be used as probes.  
 CC  
 SQ Sequence 5520 BP; 1297 A; 1491 C; 1587 G; 1145 T; 0 other;

Query Match 52.0%; Score 13; DB 19; Length 5520;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 cctcgtaagtgac 23  
 |||||||||||||  
 Db 4707 CCTGTCAGGTGC 4695

RESULT 37  
 AAF25833/C  
 ID AAF25833 standard; DNA: 130480 BP.  
 XX  
 AC AAF25833;  
 XX  
 DT 12-APR-2001 (first entry)  
 XX  
 DE R. marinus bacteriophage RM387 genomic DNA.  
 XX  
 KW Genome; thermophilic enzyme; washing powder; bleaching; ds.  
 XX  
 OS Rhodothermus marinus.  
 XX  
 PN W0200075335-A2.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PF 02-JUN-2000; 2000WO-IB00893.  
 XX  
 PR 02-JUN-1999; 99US-0137120.  
 XX  
 PA (DECO-) DECODE GENETICS EHF.  
 XX  
 PI Hjortelidsdottir S, Hreggvidsson GO, Fridjonsson OH, Aevarrsson A;  
 PI Kristjansson JK;  
 XX  
 DR WPI: 2001-061727/07.  
 XX  
 PT Isolated nucleic acid from bacteriophage RM 378 encoding enzymes useful  
 PT in recombinant DNA technology -  
 XX  
 PS Claim 3; Fig 1A-1Q2; 138pp; English.  
 XX  
 CC This invention describes a novel isolated nucleic molecule (I) comprising  
 CC the genome of bacteriophage RM 378. The invention also describes (1) an  
 CC isolated nucleic acid which encodes a polypeptide obtainable from  
 CC bacteriophage RM 378, or its active derivative or fragment; (2) an  
 CC isolated bacteriophage RM 378 (II); (3) a DNA construct (III) comprising  
 CC operatively linked to a regulatory sequence; (4) a host cell comprising

CC (III); and (5) an isolated polypeptide (IV) obtainable from (II), its  
 CC active derivative or fragment. Bacteriophage RM 378 is useful for  
 CC producing thermophilic enzymes useful in DNA research and commercial  
 CC settings (e.g. proteases and lipases used in washing powder, hydrolytic  
 CC enzymes used in bleaching). The isolated nucleic acid molecules and  
 CC vectors are useful in the manufacture of encoded polypeptide, as probes  
 CC for isolating homologous sequences (e.g. from other bacteriophage  
 CC species), as well as for detecting the presence of the bacteriophage in  
 CC a culture of host cells. The polypeptides can be used as a molecular  
 CC weight marker on SDS-PAGE gels or on molecular sieve gel filtration  
 CC columns. Because the host organism of the RM378 bacteriophage is a  
 CC thermophile, the enzymes and proteins of the RM378 bacteriophage are  
 CC significantly more thermostable than those of other (e.g. mesophilic)  
 CC bacteriophages, such as the T4 bacteriophage of *Escherichia coli*. The  
 CC enhanced stability of the enzymes and proteins of RM378 bacteriophage  
 CC allows their use under temperature conditions which would be prohibitive  
 CC for other enzymes, thus increasing the range of conditions which can be  
 CC employed not only in DNA research but also in commercial settings.

SO Sequence 130480 BP; 37304 A; 27752 C; 27687 G; 37737 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 130480;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcattcgatcct 13  
 |||||  
 Db 115871 GCATTTCGATCCT 115859

## RESULT 38

AAZ01425  
 ID AAZ01425 standard; DNA; 1038602 BP.

XX AAZ01425;

XX 07-OCT-1999 (first entry)

DE Complete genome sequence of *Chlamydia trachomatis*.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perilepitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumonia; venereal lymphogranulomatosis; ss.

XX *Chlamydia trachomatis*.

OS WO928475-A2.

PN 10-JUN-1999.

PD 27-NOV-1998; 98WO-1B01939.

XX 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997; 97FR-0015041.

PR 17-DEC-1997; 97FR-0016034.

XX (GEST ) GENSET.

PI Griffiths R;

DR WPI; 1999-371125/31.

XX Genome sequence of *Chlamydia trachomatis*

CC The present sequence represents the complete genome of *Chlamydia*

CC trachomatis. Open reading frames (ORFs) of the genome encode

CC polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines

CC against *Chlamydia trachomatis*. Antisense and ribozyme sequences can also

CC responsible for a large number of diseases, e.g. eye diseases such as  
 CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion  
 CC conjunctivitis; genital diseases such as nongonococcal urethritis;  
 CC epididymitis, cervicitis, salpingitis, perilepitis, bartholinitis;  
 CC pneumonia in breast feeding infants; and venereal  
 CC lymphogranulomatosis. The polypeptides of the invention may be of use in  
 CC treating these diseases.

SO Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 52.0%; Score 13; DB 20; Length 1038602;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ctctcaggtgca 24  
 |||||  
 Db 405349 ctctcaggtgca 405361

## RESULT 39

AAQ51957/C  
 ID AAQ51957 standard; RNA; 19 BP.

XX AAQ51957;

XX 26-MAY-1994 (first entry)

DE BCL-2 mRNA ribozyme cleavable nucleotide (1807).

XX Multiple drug resistance; mdr-1; ribozyme; membrane protein; liver;  
 KW resistance; chemotherapeutic agent; colchicine; doxorubicin; colon;  
 KW actinomycin D; vinblastine; small intestine; kidney; adrenal gland;  
 KW adenocarcinoma; bowel; transformed phenotype; promyelocytic leukemia;  
 KW human; chronic myelogenous leukemia; CML; follicular lymphoma;  
 KW B-cell acute lymphocytic leukemia; breast cancer; colon carcinoma;  
 KW neuroblastoma; lung cancer; genetic drift; mutation; hammerhead motif;  
 KW hairpin; hepatitis delta virus; group I intron; RNasep; ss.

OS Homo sapiens.

PN WO9323057-A.

XX 25-NOV-1993.

XX 13-MAY-1993; 93WO-US04573.

XX 14-MAY-1992; 92US-0882822.

PR 14-MAY-1992; 92US-0882885.

PR 26-AUG-1992; 92US-0936110.

PR 26-AUG-1992; 92US-0936421.

PR 26-AUG-1992; 92US-0936422.

PR 26-AUG-1992; 92US-0936531.

PR 07-DEC-1992; 92US-0987131.

PR 19-JAN-1993; 93US-0006122.

PR 19-JAN-1993; 93US-0008910.

XX (RIBO-) RIBOZYME PHARM INC.

XX Draper KG, Thompson JD;

XX WPI; 1993-386203/48.

XX New enzymatic RNA molecules (ribozymes) - which cleave mRNA

XX associated with tumors or mRNA expressed from gene encoding

XX multiple drug resistance

XX Claim 3; Fig 6; 69pp; English.

CC The sequences given in AAQ51825-2266 represent areas of mRNAs which are

CC associated with development or maintenance of chronic myelogenous

CC leukemia (CML), promyelocytic leukemia, Burkitt's lymphoma, or



CC acute lymphocytic leukemia, follicular lymphoma, B-cell acute  
 CC lymphocytic leukemia, breast cancer, colon carcinoma, neuroblastoma  
 CC and lung cancer. The full length mRNAs containing these target  
 CC sequences, encode aberrant cellular proteins which are able to control  
 CC cellular proliferation and are directly linked to a leukemic  
 CC phenotype. These target sequences are identified by the ribozyme of  
 CC the invention. The ribozymes is formed in a hammerhead motif, but may  
 CC also be formed in the motif of a hairpin, hepatitis delta virus, group  
 CC I intron or RNaseP-like RNA. These ribozymes may be used to inhibit  
 CC the development or expression of a transformed phenotype in man and  
 CC other animals by modulating expression of the corresponding gene.  
 CC Cleavage of target mRNAs expressed in pre-neoplastic and transformed  
 CC cells elicits inhibition of the transformed state. Multiple drug  
 CC resistance (mdr-1) mRNA specific ribozymes remove the mechanism of  
 CC drug resistance used by transformed cells and thus enhances drug  
 CC therapies for tumours. The ribozymes may also be used to study  
 CC genetic drift and mutations within cells.  
 XX  
 SO Sequence 19 BP; 4 A; 8 C; 5 G; 2 U; 0 other;

Query Match 48.0%; Score 12; DB 14; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 3,7e+02;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 gctcagtgctgag 25  
 |||||||||  
 Db 18 GCTCAGTCTCAG 7

RESULT 40  
 AAA66407  
 ID AAA66407 standard; DNA; 20 BP.  
 XX  
 AC AAA66407;  
 DT 09-OCT-2000 (first entry)  
 XX  
 DE Dog genomic marker oligonucleotide sequence SEQ ID NO:269.  
 XX  
 XX Dog; genome; genomic marker; radiation hybrid map; identification;  
 KM chromosome location; gene marker; polymorphic microsatellite marker;  
 KM phenotype; behaviour; pedigree; ss.  
 XX  
 OS Canis familiaris.  
 XX  
 PN WO200029615-A2.  
 XX  
 PD 25-MAY-2000.  
 XX  
 PF 15-NOV-1999; 99WO-IB01907.  
 XX  
 PR 13-NOV-1998; 98US-0108193.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Galibert F, Andre C;  
 DR WPI: 2000-387821/33.  
 XX  
 PT New radiation hybrid map of the dog, Canine familiaris, genome, useful  
 PT for e.g. identifying genes implicated in phenotypic and behavioral  
 PT traits or in genetic diseases and for studying dog pedigrees -  
 XX  
 XX Claim 1; Page 64; 87pp; English.  
 PS  
 CC The present invention describes a radiation hybrid map of the dog  
 CC (Canine familiaris) genome comprising the genome location of a marker  
 CC selected from AAA66139 to AAA66942. The radiation hybrid map is useful  
 CC for identifying and localising dog genes, since it covers approximately  
 CC 80 % of the dog genome and provides a dense map integrating different  
 CC types (i.e. Type I and Type II) of markers. The map and the dog genome  
 CC markers (or complementary sequences) are especially useful to identify

CC genes responsible for phenotypic and behavioural traits in dogs, to  
 CC identify morbid genes, to analyse diseases and identify implicated genes  
 CC in such diseases and their alleles, and to study dog pedigrees. They  
 CC may also be useful for isolating corresponding human gene sequences  
 CC e.g. genes involved in genetic diseases.  
 XX  
 SO Sequence 20 BP; 2 A; 8 C; 3 G; 7 T; 0 other;

Query Match 48.0%; Score 12; DB 21; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 3,7e+02;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gctcctgctgag 19  
 |||||||||  
 Db 7 gctcctgctgag 18

RESULT 41  
 AA050268/C  
 ID AA050268 standard; DNA; 50 BP.  
 XX  
 AC AA050268;  
 DT 06-MAY-1994 (first entry)  
 XX  
 DE HIV env INS mutagenic oligonucleotide M24 (8072-8121).  
 XX  
 KM HIV-1; p17gag; inhibition; gag; M1; M2; M3; M4; silent mutation; mRNA;  
 KM vector; p17; point mutation; p17M1234; Hlat cells; gene replacement;  
 KM inhibitor/instability signal; INS; stability; utilisation; vaccine;  
 KM interferon; interleukin; fos proto-oncogene protein; growth factor;  
 KM env; attenuated; AIDS; AIDS-related disease; latent infection;  
 KM gene therapy; Human immunodeficiency virus type 1; p24;  
 KM Long terminal repeat; LTR; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9320212-A.  
 XX  
 PD 14-OCT-1993.  
 XX  
 PF 29-MAR-1993; 93WO-US02908.  
 XX  
 PR 27-MAR-1992; 92US-0858747.  
 XX  
 PA (USSH ) US SEC DEPT HEALTH.  
 XX  
 PI Felder BK, Pavlakis GN;  
 DR WPI: 1993-336919/42.  
 XX  
 PT Eliminating inhibitory-instability regions in mRNA to improve  
 PT stability and expression - by making multiple point mutations  
 PT within A-T rich regions  
 XX  
 XX Example 4; Page 80; 117pp; English.  
 PS  
 CC The sequences given in AA050202-70 are oligonucleotides which were  
 CC used to illustrate the method of the invention for the mutagenesis of  
 CC inhibitory/instability signals (INS). Mutation in an INS produced  
 CC using oligonucleotides such as these, increases the stability and/or  
 CC utilisation of mRNA without changing its protein coding capacity, or  
 CC if the sequence is changed, its function is maintained. Other genes  
 CC encoding such mRNA molecules include growth factor, interferon,  
 CC interleukin, fos proto-oncogene protein and HIV env and gag gene  
 CC proteins. Nucleic acid constructs in which INS function has been  
 CC impeded, can be used as vaccines, esp. against AIDS and AIDS-related  
 CC diseases by preventing HIV from establishing a latent infection, as  
 CC is possible using the INS, and thus escaping immune system  
 CC surveillance. The constructs may also be used in gene therapy for  
 CC gene replacement by homologous recombination with a target gene in  
 CC situ. See also AA050200-02.



Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 gatctctcag 19  
 |||||  
 Db 37 gatctctcag 48

RESULT 44  
 AAX12922/c  
 ID AAX12922 standard; DNA: 110 BP.  
 XX

AC AAX12922;

DT 30-MAR-1999 (first entry)

DE Human biallelic polymorphic DNA fragment EST530186.

KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
 detection; phenotypic typing; characteristic; infection; hereditary;  
 autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
 treatment; marker; ss.

XX Homo sapiens.

OS WO9820165-A2.

PD 14-MAY-1998.

PF 05-NOV-1997; 97WO-US20313.

PR 06-NOV-1996; 96US-0030455.

PA (WHEED) WHITEHEAD INST BIOMEDICAL RES.

PI Hudson T, Lander ES, Wang D;

DR WPI: 1998-286974/25.

PT New isolated nucleic acid segments from the human genome - used for  
 PT determining polymorphic forms for use in e.g. forensics, paternity  
 PT testing or phenotypic typing for disease

PS Claim 1; Page 303; 310pp; English.

XX AAX10269-X12937 are human DNA fragments which contain biallelic  
 CC polymorphic markers which have been isolated using the primers  
 CC represented in AAX09121-X10268. The base occupying the polymorphic site  
 CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments  
 CC can be used in methods for determining polymorphic forms in an individual  
 CC for use in e.g. forensics, paternity testing or for phenotypic typing for  
 CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan  
 CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,  
 CC familial hypercholesterolemia, polycystic kidney disease, hereditary  
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
 CC system, infection by pathogenic microorganisms, and characteristics such  
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
 CC endurance, fertility, and susceptibility or receptivity to particular  
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
 CC segments can also be used to produce medicaments for the treatment or  
 CC prophylaxis of such diseases.

XX Sequence 110 BP; 33 A; 21 C; 37 G; 19 T; 0 other;

Query Match 48.0%; Score 12; DB 19; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 ttgatctcgt 16  
 |||||

Db 98 TTCGATCTCGT 87

RESULT 45  
 AAX11656/c  
 ID AAX11656 standard; DNA: 110 BP.  
 XX

AC AAX11656;

DT 30-MAR-1999 (first entry)

DE Human biallelic polymorphic DNA fragment EST530186.

KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
 detection; phenotypic typing; characteristic; infection; hereditary;  
 autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
 treatment; marker; ss.

XX Homo sapiens.

OS WO9820165-A2.

PD 14-MAY-1998.

PF 05-NOV-1997; 97WO-US20313.

PR 06-NOV-1996; 96US-0030455.

PA (WHEED) WHITEHEAD INST BIOMEDICAL RES.

PI Hudson T, Lander ES, Wang D;

DR WPI: 1998-286974/25.

PT New isolated nucleic acid segments from the human genome - used for  
 PT determining polymorphic forms for use in e.g. forensics, paternity  
 PT testing or phenotypic typing for disease

PS Claim 1; Page 179; 310pp; English.

XX AAX10269-X12937 are human DNA fragments which contain biallelic  
 CC polymorphic markers which have been isolated using the primers  
 CC represented in AAX09121-X10268. The base occupying the polymorphic site  
 CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments  
 CC can be used in methods for determining polymorphic forms in an individual  
 CC for use in e.g. forensics, paternity testing or for phenotypic typing for  
 CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan  
 CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,  
 CC familial hypercholesterolemia, polycystic kidney disease, hereditary  
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
 CC system, infection by pathogenic microorganisms, and characteristics such  
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
 CC endurance, fertility, and susceptibility or receptivity to particular  
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
 CC segments can also be used to produce medicaments for the treatment or  
 CC prophylaxis of such diseases.

XX Sequence 110 BP; 32 A; 21 C; 37 G; 19 T; 1 other;

Query Match 48.0%; Score 12; DB 19; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 ttgatctcgt 16  
 |||||  
 Db 98 TTCGATCTCTGCT 87

Wed Oct 10 07:46:08 2001

us-09-396-196f-8.oli.rng

Page 34

Search completed: October 9, 2001, 15:53:10  
Job time: 15151 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:39 ; Search time 218.82 Seconds  
(Without alignments)  
21.629 Million cell updates/sec

Title: US-09-396-196F-8

Sequence: 1 gatttcgacccctcgtcagtgag 25

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 9465562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 8847

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCUTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description       |
|------------|-------|---------|--------------|-------|-------------------|
| 1          | 25    | 100.0   | 1041         | 2     | US-08-401-068-7   |
| 2          | 25    | 100.0   | 1041         | 2     | US-08-846-338-7   |
| 3          | 25    | 100.0   | 5872         | 3     | US-08-411-768B-1  |
| 4          | 25    | 100.0   | 5872         | 3     | US-08-411-768B-6  |
| 5          | 14    | 56.0    | 66           | 3     | US-09-023-228B-96 |
| 6          | 14    | 56.0    | 2104         | 1     | US-08-682-193A-1  |
| 7          | 12    | 48.0    | 19           | 1     | US-07-936-421-11  |
| 8          | 12    | 48.0    | 50           | 2     | US-08-850-049-122 |
| 9          | 12    | 48.0    | 50           | 2     | US-08-050-478-122 |
| 10         | 12    | 48.0    | 61           | 4     | US-08-687-421-385 |
| 11         | 12    | 48.0    | 500          | 3     | US-09-248-528-1   |
| 12         | 12    | 48.0    | 500          | 4     | US-09-549-108-1   |
| 13         | 12    | 48.0    | 500          | 4     | US-09-549-111-1   |
| 14         | 12    | 48.0    | 500          | 4     | US-09-549-106-1   |
| 15         | 12    | 48.0    | 615          | 2     | US-08-465-485A-22 |
| 16         | 12    | 48.0    | 615          | 3     | US-09-080-285-22  |
| 17         | 12    | 48.0    | 623          | 6     | 5506344-3         |
| 18         | 12    | 48.0    | 717          | 2     | US-08-465-485A-20 |
| 19         | 12    | 48.0    | 717          | 3     | US-09-080-285-20  |
| 20         | 12    | 48.0    | 760          | 1     | US-08-405-702A-11 |
| 21         | 12    | 48.0    | 780          | 3     | US-08-927-128-3   |
| 22         | 12    | 48.0    | 831          | 6     | 5459251-3         |
| 23         | 12    | 48.0    | 831          | 6     | 5506344-4         |
| 24         | 12    | 48.0    | 864          | 2     | US-08-595-868C-11 |
| 25         | 12    | 48.0    | 867          | 4     | US-09-139-819A-11 |
| 26         | 12    | 48.0    | 876          | 1     | US-08-726-725-5   |
| 27         | 12    | 48.0    | 911          | 4     | US-09-126-109-5   |

|      |    |      |      |   |                   |                    |
|------|----|------|------|---|-------------------|--------------------|
| C 28 | 12 | 48.0 | 911  | 5 | PCT-US93-06251-3  | Sequence 3, Appl1  |
| 29   | 12 | 48.0 | 1411 | 1 | US-07-952-817-10  | Sequence 10, Appl1 |
| 30   | 12 | 48.0 | 1411 | 6 | 5210025-3         | Patent No. 5210025 |
| 31   | 12 | 48.0 | 1516 | 3 | US-09-248-528-3   | Sequence 3, Appl1  |
| 32   | 12 | 48.0 | 1516 | 4 | US-09-549-108-3   | Sequence 3, Appl1  |
| 33   | 12 | 48.0 | 1516 | 4 | US-09-549-111-3   | Sequence 3, Appl1  |
| 34   | 12 | 48.0 | 1516 | 4 | US-09-549-106-3   | Sequence 3, Appl1  |
| 35   | 12 | 48.0 | 1517 | 3 | US-09-248-528-2   | Sequence 2, Appl1  |
| 36   | 12 | 48.0 | 1517 | 4 | US-09-549-108-2   | Sequence 2, Appl1  |
| 37   | 12 | 48.0 | 1517 | 4 | US-09-549-111-2   | Sequence 2, Appl1  |
| 38   | 12 | 48.0 | 1517 | 4 | US-09-549-106-2   | Sequence 2, Appl1  |
| 39   | 12 | 48.0 | 1602 | 4 | US-08-987-151-4   | Sequence 4, Appl1  |
| 40   | 12 | 48.0 | 1785 | 4 | US-08-987-151-1   | Sequence 4, Appl1  |
| 41   | 12 | 48.0 | 1785 | 4 | US-08-987-151-3   | Sequence 3, Appl1  |
| 42   | 12 | 48.0 | 1846 | 2 | US-08-365-486A-16 | Sequence 16, Appl1 |
| 43   | 12 | 48.0 | 1846 | 4 | US-08-880-342-16  | Sequence 16, Appl1 |
| 44   | 12 | 48.0 | 1889 | 2 | US-09-026-587-2   | Sequence 2, Appl1  |
| 45   | 12 | 48.0 | 1889 | 2 | US-09-227-420-2   | Sequence 2, Appl1  |

#### ALIGNMENTS

RESULT 1  
US-08-401-068-7  
Sequence 7, Application US/08401068  
Patent No. 5859335  
GENERAL INFORMATION:  
APPLICANT: Patton, David  
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESS: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/401,068  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/351,970  
FILING DATE: 08-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1041 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1038  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /product= "Biotin synthase"  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
US-08-401-068-7



```

LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORP1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob_RBS_no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc_RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioA_RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1_RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent"
OTHER INFORMATION: /transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match      100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 7.3e+06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY      1 gcattgcattcgtcgatgagcag 25
        |||
DB       215 GCATTGCATTCCTGCGACAGTGCAG 239

RESULT      4
US-08-411-768B-6
: Sequence 6, Application US/08411768B
: Patient No. 6083712
: GENERAL INFORMATION:
: APPLICANT: Olwen Birch
: APPLICANT: Johann Brass
:
```

APPLICANT: Martin Fuhrmann  
 APPLICANT: Nicholas Shaw  
 TITLE OF INVENTION: Biotechnological Method  
 TITLE OF INVENTION: of Producing Biotin  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
 STREET: 30 Rockefeller Plaza  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10112  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Wordperfect  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/411,768B  
 FILING DATE: 31-March-95  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: CH 3124/92  
 FILING DATE: 02-OCT-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: CH 2134/93  
 FILING DATE: 15-JUL-1993  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5872 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Escherichia coli  
 STRAIN: DSM498  
 IMMEDIATE SOURCE:  
 CLONE: pBO30A15-9  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1154..2308  
 IDENTIFICATION METHOD: experimental  
 OTHER INFORMATION: /codon\_start=1154  
 OTHER INFORMATION: /BC\_number=2.3.1.47  
 OTHER INFORMATION: /product="KAPA synthase"  
 OTHER INFORMATION: /evidence=EXPERIMENTAL  
 OTHER INFORMATION: /gene="bioD"  
 OTHER INFORMATION: /number=2  
 OTHER INFORMATION: /standard\_name="8-Amino-7-oxononanoate synthase"  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 3043..3753  
 IDENTIFICATION METHOD: experimental  
 OTHER INFORMATION: /codon\_start=3043  
 OTHER INFORMATION: /BC\_number=6.3.3.3  
 OTHER INFORMATION: /product="DTB synthase"  
 OTHER INFORMATION: /evidence=EXPERIMENTAL  
 OTHER INFORMATION: /gene="bioD"  
 OTHER INFORMATION: /number=4  
 OTHER INFORMATION: /standard\_name="Dethiobiotin synthase"  
 FEATURE:  
 NAME/KEY: RBS  
 LOCATION: 1141..1156  
 OTHER INFORMATION: /standard\_name="biof RBS"  
 FEATURE:  
 NAME/KEY: RBS  
 LOCATION: 3030..3045  
 OTHER INFORMATION: /standard\_name="biof RBS"  
 PUBLICATION INFORMATION:  
 DOCUMENT NUMBER: WO 87/01391 B1

;; FILING DATE: 26-AUG-1986  
;; PUBLICATION DATE: 07-APR-1993  
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;  
Best Local Similarity 100.0%; Pred. No. 7.3e-06;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgcctcgtcaglgcag 25  
|||||  
DB 215 GCATTCGATCCTCGTCAGTGACG 239

RESULT 5  
US-09-023-228B-96  
; Sequence 96, Application US/09023228B  
; Patent No. 6140490  
; GENERAL INFORMATION:  
; APPLICANT: BIESECKER, GREGORY  
; APPLICANT: GOLD, LARRY  
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF  
; TITLE OF INVENTION: COMPLEMENT SYSTEM PROTEINS  
; NUMBER OF SEQUENCES: 157  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Swanson & Bratschun, L.L.C.  
; STREET: 8400 E. Prentice Place #200  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: US  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 8.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,228B  
; FILING DATE: 12-FEBRUARY-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/01739  
; FILING DATE: 30 JAN 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/595,335  
; FILING DATE: 1 FEB 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barry J. Swanson, Esq.  
; REGISTRATION NUMBER: 33,215  
; REFERENCE/DOCKET NUMBER: NEX50/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 793-3333  
; TELEFAX: (303) 793-3433  
; INFORMATION FOR SEQ ID NO: 96:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 66 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: RNA  
; FEATURE:  
; OTHER INFORMATION: All pyrimidines are 2'-F modified  
US-09-023-228B-96

Query Match 56.0%; Score 14; DB 3; Length 66;  
Best Local Similarity 71.4%; Pred. No. 13;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 6 tcgactcctcgtcag 19  
:|||||  
DB 40 UCGAUCUCUCUCAG 53

RESULT 6  
US-08-682-193A-1  
; Sequence 1, Application US/08682193A  
; Patent No. 5776740  
; GENERAL INFORMATION:  
; APPLICANT: HATAKEYAMA, Kazuhisa  
; APPLICANT: GOTO, Makoto  
; APPLICANT: TERASAWA, Masato  
; APPLICANT: YUKAWA, Hideaki  
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF  
; TITLE OF INVENTION: L-TRYPTOPHAN  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
; STREET: 130 WATER STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02019  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/682,193A  
; FILING DATE: 17-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 7-181730  
; FILING DATE: 18-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: DAVID, RESNICK S  
; REGISTRATION NUMBER: 34,235  
; REFERENCE/DOCKET NUMBER: 46643  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-523-3400  
; TELEFAX: 617-523-6440  
; TELEX: 200291 STRB  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2104 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; ORGANISM: Brevibacterium flavum  
; STRAIN: MJ-233  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 556...1855  
; OTHER INFORMATION:  
US-08-682-193A-1

Query Match 56.0%; Score 14; DB 1; Length 2104;  
Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttgcgtcctcgtc 17  
|||||  
DB 1635 TTTCGATCCTCGTC 1648

RESULT 7  
US-07-936-421-11/C  
; Sequence 11, Application US/07936421  
; Patent No. 5750390  
; GENERAL INFORMATION:



APPLICANT: James D. Thompson  
APPLICANT: Kenneth G. Draper  
TITLE OF INVENTION: METHOD AND REAGENT FOR  
TITLE OF INVENTION: TREATMENT OF DISEASES CAUSED  
TITLE OF INVENTION: BY EXPRESSION OF THE BCL-2  
TITLE OF INVENTION: GENE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 611 West Sixth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/936,421  
FILING DATE: 19920826  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: none  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 197/243  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ. ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-936-421-11  
Query Match 48.0%; Score 12; DB 1; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 14 cgtcaggtgcag 25  
DB 18 CGTCAGGTGCAG 7  
RESULT 8  
US-08-850-049-122/C  
Sequence 122, Application US/08850049  
Patent No. 5965726  
GENERAL INFORMATION:  
APPLICANT:  
APPLICANT:  
APPLICANT:  
TITLE OF INVENTION: METHOD OF ELIMINATING  
TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF mRNA  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM P.C. DOS (Version 5.0)  
OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/936,421  
FILING DATE: 19920826  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: none  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 197/243  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ. ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-936-421-11  
Query Match 48.0%; Score 12; DB 1; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/850,049  
FILING DATE: 02-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/050,478  
FILING DATE: 26-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/02908  
FILING DATE: 29-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/858,747  
FILING DATE: 27-MAR-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MORRIS, MARY J.  
REGISTRATION NUMBER: 34,398  
REFERENCE/DOCKET NUMBER: 2026-4006US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)758-4800  
TELEFAX: (212)751-6849  
INFORMATION FOR SEQ. ID NO: 122:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
US-08-850-049-122  
Query Match 48.0%; Score 12; DB 2; Length 50;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 cgatcctcgta 18  
DB 43 CGATCCTCGTCA 32  
RESULT 9  
US-08-050-478-122/C  
Sequence 122, Application US/08050478  
Patent No. 5972596  
GENERAL INFORMATION:  
APPLICANT:  
APPLICANT:  
APPLICANT:  
TITLE OF INVENTION: METHOD OF ELIMINATING  
TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF mRNA  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/050,478  
FILING DATE: 26-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/02908  
FILING DATE: 29-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/858,747  
FILING DATE: 27-MAR-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MORRIS, MARY J.  
REGISTRATION NUMBER: 34,398  
REFERENCE/DOCKET NUMBER: 2026-4006US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)751-6849  
TELEFAX: (212)751-6849  
INFORMATION FOR SEQ ID NO: 122:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
US-08-050-478-122

Query Match 48.0%; Score 12; DB 2; Length 50;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 cgatccctcgta 18  
|||||  
DB 43 CGATCCTCGTCA 32

RESULT 10  
US-08-687-421-385  
Sequence 385, Application US/08687421  
Patent No. 6177557  
GENERAL INFORMATION:  
APPLICANT: Gold, Larry  
APPLICANT: Janjic, Nebojsa  
APPLICANT: Tassel, Diane  
TITLE OF INVENTION: HIGH-AFFINITY LIGANDS OF BASIC  
TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR AND  
NUMBER OF SEQUENCES: 445  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Swanson & Bratschun, L.L.C.  
STREET: 8400 E. Prentice Avenue, Suite 200  
CITY: Englewood  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,421  
FILING DATE: 08-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/195,005  
FILING DATE: 10-FEBRUARY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE: 22-APRIL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/219,012  
FILING DATE: 28-MARCH-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,333  
FILING DATE: 11-NOVEMBER-1992  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/714,131  
FILING DATE: 10-JUNE-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/536,428  
FILING DATE: 11-JUNE-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Barry J. Swanson  
REGISTRATION NUMBER: 33,215  
REFERENCE/DOCKET NUMBER: NEX07/PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 793-3333  
TELEFAX: (303) 793-3433  
INFORMATION FOR SEQ ID NO: 385:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 61 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-687-421-385

Query Match 48.0%; Score 12; DB 4; Length 61;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gatccctcgta 19  
|||||  
DB 37 GATCCTCGTCA 48

RESULT 11  
US-09-248-528-1/c  
Sequence 1, Application US/09248528C  
Patent No. 6153415  
GENERAL INFORMATION:  
APPLICANT: Oriol, Patrick J  
APPLICANT: Padmakumar, Rugmini  
APPLICANT: Kim, Sang H  
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite  
TITLE OF INVENTION: Hydrazide from a Thermophilic Bacillus  
FILE REFERENCE: MSQ 4.1-401  
CURRENT APPLICATION NUMBER: US/09/248,528C  
CURRENT FILING DATE: 1999-02-10  
EARLIER APPLICATION NUMBER: 60/083,485  
EARLIER FILING DATE: 1998-04-29  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 500  
TYPE: DNA  
ORGANISM: BR449  
FEATURE:  
NAME/KEY: rRNA  
LOCATION: (1)..(500)  
US-09-248-528-1

Query Match 48.0%; Score 12; DB 3; Length 500;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tccctcgta 21  
|||||  
DB 484 TCCTCGTCA 473

RESULT 12  
US-09-549-108-1/c  
Sequence 1, Application US/09549108  
Patent No. 6214603  
GENERAL INFORMATION:  
APPLICANT: Oriol, Patrick J  
APPLICANT: Padmakumar, Rugmini

APPLICANT: Kim, Sang H  
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile  
TITLE OF INVENTION: Hydrolase from a Thermophilic Bacillus  
FILE REFERENCE: MSU 4.1-486  
CURRENT APPLICATION NUMBER: US/09/549,108  
CURRENT FILING DATE: 2000-04-13  
PRIOR APPLICATION NUMBER: 60/083,485  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 09/248,528  
PRIOR FILING DATE: 1999-02-10  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 500  
TYPE: DNA  
ORGANISM: BR449  
FEATURE:  
NAME/KEY: rRNA  
LOCATION: (1)..(500)  
US-09-549-108-1

Query Match 48.0%; Score 12; DB 4; Length 500;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctctgcagct 21  
|||||  
DB 484 TCCTCGTCAGCT 473

RESULT 13  
US-09-549-111-1/c  
Sequence 1, Application US/09549111  
Patent No. 6228633  
GENERAL INFORMATION:  
APPLICANT: Orieli, Patrick J  
APPLICANT: Padmakumar, Rugmini  
APPLICANT: Kim, Sang H  
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile  
TITLE OF INVENTION: Hydrolase from a Thermophilic Bacillus  
FILE REFERENCE: MSU 4.1-489  
CURRENT APPLICATION NUMBER: US/09/549,111  
CURRENT FILING DATE: 2000-04-13  
PRIOR APPLICATION NUMBER: 60/083,485  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 09/248,528  
PRIOR FILING DATE: 1999-02-10  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 500  
TYPE: DNA  
ORGANISM: BR449  
FEATURE:  
NAME/KEY: rRNA  
LOCATION: (1)..(500)  
US-09-549-111-1

Query Match 48.0%; Score 12; DB 4; Length 500;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctctgcagct 21  
|||||  
DB 484 TCCTCGTCAGCT 473

RESULT 14  
US-09-549-106-1/c  
Sequence 1, Application US/09549106  
Patent No. 6242242

GENERAL INFORMATION:  
APPLICANT: Orieli, Patrick J  
APPLICANT: Padmakumar, Rugmini  
APPLICANT: Kim, Sang H  
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile  
TITLE OF INVENTION: Hydrolase from a Thermophilic Bacillus  
FILE REFERENCE: MSU 4.1-487  
CURRENT APPLICATION NUMBER: US/09/549,106  
CURRENT FILING DATE: 2000-04-13  
PRIOR APPLICATION NUMBER: 60/083,485  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 09/248,528  
PRIOR FILING DATE: 1999-02-10  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 500  
TYPE: DNA  
ORGANISM: BR449  
FEATURE:  
NAME/KEY: rRNA  
LOCATION: (1)..(500)  
US-09-549-106-1

Query Match 48.0%; Score 12; DB 4; Length 500;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctctgcagct 21  
|||||  
DB 484 TCCTCGTCAGCT 473

RESULT 15  
US-08-465-485A-22/c  
Sequence 22, Application US/08465485A  
Patent No. 5831066  
GENERAL INFORMATION:  
APPLICANT: Reed, John  
TITLE OF INVENTION: Regulation of bcl-2 Gene Expression  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1735 S. Jefferson Davis Hwy., Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,485A  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/124,256  
FILING DATE: 20-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/840,716  
FILING DATE: 21-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/288,692  
FILING DATE: 22-DEC-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Fortney, Andrew D.  
REGISTRATION NUMBER: 34,600  
REFERENCE/DOCKET NUMBER: 3335-070-55 CONT  
TELECOMMUNICATION INFORMATION:

```

:
: TELEPHONE: (408) 436-2070
: TELEFAX: (408) 436-2075
: INFORMATION FOR SEQ ID NO: 22:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 615 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 1..615
: US-08-465-485A-22
:
Query Match          48.0%; Score 12; DB 2; Length 615;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 cgtcaggtgcag 25
Db 366 CGTCAGGTGCAG 355

RESULT 16
US-09-080-285-22/c
: Sequence 22; Application US/09080285
: Patent No. 6040181
: GENERAL INFORMATION:
:   APPLICANT: Reed, John
:   TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
:   NUMBER OF SEQUENCES: 29
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
:     STREET: 1755 S. Jefferson Davis Hwy., Suite 400
:     CITY: Arlington
:     STATE: Virginia
:     COUNTRY: U.S.A.
:     ZIP: 22202
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentin Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/09/080,285
:     FILING DATE:
:   CLASSIFICATION:
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: US 08/465,485
:       FILING DATE: 05-JUN-1995
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER:
:       FILING DATE: 20-SEP-1993
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: US 07/840,716
:       FILING DATE: 21-FEB-1992
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: US 07/288,692
:       FILING DATE: 22-DEC-1988
:     ATTORNEY/AGENT INFORMATION:
:       NAME: Fortney, Andrew D.
:       REGISTRATION NUMBER: 34,600
:       REFERENCE/DOCKET NUMBER: 3335-070-55 CONT
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: (408) 436-2070
:       TELEFAX: (408) 436-2075
:     INFORMATION FOR SEQ ID NO: 22:
:       SEQUENCE CHARACTERISTICS:
:         LENGTH: 615 base pairs
:         TYPE: nucleic acid
:         STRANDEDNESS: single

```

```

:
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 1..615
: US-09-080-285-22
:
Query Match          48.0%; Score 12; DB 3; Length 615;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 cgtcaggtgcag 25
Db 366 CGTCAGGTGCAG 355

RESULT 17
5506344-3/c
: Patent No. 5506344
: APPLICANT: TSUJIMOTO, YOSHIIHIDE, CROCE, CARLO A.
:   TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR BCL-2 GENE PRODUCT
:   NUMBER OF SEQUENCES: 5
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/435,193
:     FILING DATE: 05-MAY-1995
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: 228,704
:       FILING DATE: 18-APR-1994
:       APPLICATION NUMBER: 994,941
:       FILING DATE: 23-DEC-1992
:       APPLICATION NUMBER: 663,010
:       FILING DATE: 19-MAR-1991
:       APPLICATION NUMBER: 883,687
:       FILING DATE: 09-JUL-1986
:   SEQ ID NO: 3;
:   LENGTH: 623
: 5506344-3
:
Query Match          48.0%; Score 12; DB 6; Length 623;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 cgtcaggtgcag 25
Db 366 CGTCAGGTGCAG 355

RESULT 18
US-08-465-485A-20/c
: Sequence 20; Application US/08465485A
: Patent No. 5831066
: GENERAL INFORMATION:
:   APPLICANT: Reed, John
:   TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
:   NUMBER OF SEQUENCES: 29
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
:     STREET: 1755 S. Jefferson Davis Hwy., Suite 400
:     CITY: Arlington
:     STATE: Virginia
:     COUNTRY: U.S.A.
:     ZIP: 22202
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentin Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/465,485A
:     FILING DATE: 05-JUN-1995

```

CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/124,256  
FILING DATE: 20-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/840,716  
FILING DATE: 21-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/288,692  
FILING DATE: 22-DEC-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Fortney, Andrew D.  
REGISTRATION NUMBER: 34,600  
REFERENCE/DOCKET NUMBER: 3335-070-55 CONT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (408) 436-2070  
TELEFAX: (408) 436-2075  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 717 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..717  
US-08-465-485a-20

Query Match 48.0%; Score 12; DB 2; Length 717;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtgcag 25  
|||||  
Db 366 CGTCAGTGCAG 355

RESULT 19  
US-09-080-285-20/c  
Sequence 20, Application US/09080285  
Patent No. 6040181  
GENERAL INFORMATION:  
APPLICANT: Reed, John  
TITLE OF INVENTION: Regulation of bcl-2 Gene Expression  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, McLELLAND, MAIER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 S. Jefferson Davis Hwy., Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/080,285  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/465,485  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/124,256  
FILING DATE: 20-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/840,716  
FILING DATE: 21-FEB-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/288,692  
FILING DATE: 22-DEC-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Fortney, Andrew D.  
REGISTRATION NUMBER: 34,600  
REFERENCE/DOCKET NUMBER: 3335-070-55 CONT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (408) 436-2070  
TELEFAX: (408) 436-2075  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 717 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..717  
US-09-080-285-20

Query Match 48.0%; Score 12; DB 3; Length 717;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtgcag 25  
|||||  
Db 366 CGTCAGTGCAG 355

RESULT 20  
US-08-405-702A-11/c  
Sequence 11, Application US/08405702A  
Patent No. 5789389  
GENERAL INFORMATION:  
APPLICANT: Tarasiewicz, Dariusz G  
APPLICANT: Schott, Brigitte  
APPLICANT: Holzmayer, Tatiana A.  
APPLICANT: Roninson, Igor B  
TITLE OF INVENTION: BCL2 Derived Genetic Elements Associated  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 10 South Wacker Drive, Suite 3000  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,702A  
FILING DATE: 17-MAR-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5789389nan, Kevin E  
REGISTRATION NUMBER: 35,303  
REFERENCE/DOCKET NUMBER: 95,332  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1234  
TELEFAX: 312-715-1234  
TELEX: 910-221-5317  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 760 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 21..740  
US-08-405-702A-11

Query Match 48.0%; Score 12; DB 1; Length 760;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtgcag 25  
|||||  
DB 386 CGTCAGGTGCAG 375

## RESULT 21

US-08-927-128-3  
Sequence 3, Application US/08927128  
Patent No. 6127150

## GENERAL INFORMATION:

APPLICANT: Coolidge, Thomas  
APPLICANT: Wagner, Fred  
APPLICANT: ven Heeke, Gino  
APPLICANT: Schuster, Sheldon  
APPLICANT: Stout, Jay  
APPLICANT: Wylie, Dwane  
TITLE OF INVENTION: PURIFICATION DIRECTED CLOSING OF PEPTIDES  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould  
STREET: 3100 No. 6127150west Center, 90 S. 7th Street  
CITY: Minneapolis  
STATE: MN  
COUNTRY: U.S.A.  
ZIP: 55402

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/927,128  
FILING DATE: 05-SEP-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/680,004  
FILING DATE: 15-JUL-1995

## ATTORNEY/AGENT INFORMATION:

NAME: Carter, Charles G  
REGISTRATION NUMBER: 35,093  
REFERENCE/DOCKET NUMBER: 8648.20SD1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612/332-5300  
TELEFAX: 612/332-9081  
TELEX:

## INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 780 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1..780  
OTHER INFORMATION:

US-08-927-128-3

Query Match 48.0%; Score 12; DB 3; Length 780;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcattcttgt 16  
|||||  
DB 532 TTGCATTCTTGT 543

## RESULT 22

5459251-3/C  
Patent No. 5459251

APPLICANT: Tsujimoto, Yoshide;Croce, Carlo A.  
TITLE OF INVENTION: DNA MOLECULES HAVING HUMAN BCL-2 GENE  
SEQUENCES

## NUMBER OF SEQUENCES: 4

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/228,704  
FILING DATE: 18-APR-1994

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 994,941  
FILING DATE: 23-DEC-1992  
APPLICATION NUMBER: 663,010  
FILING DATE: 19-MAR-1991  
APPLICATION NUMBER: 883,687  
FILING DATE: 09-JUL-1986

SEQ ID NO:3  
LENGTH: 831  
5459251-3

Query Match 48.0%; Score 12; DB 6; Length 831;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtgcag 25  
|||||  
DB 512 CGTCAGGTGCAG 501

## RESULT 23

5506344-4/C  
Patent No. 5506344

APPLICANT: TSUJIMOTO, YOSHIDE;CROCE, CARLO A.  
TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR BCL-2GENE PRODUCT  
NUMBER OF SEQUENCES: 5  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/435,193  
FILING DATE: 05-MAY-1995  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 228,704  
FILING DATE: 18-APR-1994  
APPLICATION NUMBER: 994,941  
FILING DATE: 23-DEC-1992  
APPLICATION NUMBER: 663,010  
FILING DATE: 19-MAR-1991  
APPLICATION NUMBER: 883,687  
FILING DATE: 09-JUL-1986

SEQ ID NO:4  
LENGTH: 831  
5506344-4

Query Match 48.0%; Score 12; DB 6; Length 831;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtgcag 25  
|||||  
DB 512 CGTCAGGTGCAG 501

RESULT 24  
US-08-595-868C-11  
; Sequence 11, Application US/08595868C  
; Patent No. 5962270  
; GENERAL INFORMATION:  
; APPLICANT: Wagner, Fred  
; APPLICANT: Stout, Jay  
; APPLICANT: Henriksen, Dennis  
; APPLICANT: Partridge, Bruce  
; APPLICANT: Holmquist, Bart  
; APPLICANT: Frank, Julie  
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 3100 No. 5962270west Center, 90 S. 7th Street  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: U.S.A.  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/595.868C  
; FILING DATE: 06-FEB-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carter, Charles G  
; REGISTRATION NUMBER: 35,093  
; REFERENCE/DOCKET NUMBER: 8648.590501  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612/332-5300  
; TELEFAX: 612/332-9081  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 864 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 1...864  
; OTHER INFORMATION:  
US-08-595-868C-11

Query Match 48.0%; Score 12; DB 2; Length 864;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttcgattcctcgt 16  
|||||

Db 532 TTCGATCCTCGT 543

RESULT 25  
US-09-139-819A-11  
; Sequence 11, Application US/09139819A  
; Patent No. 6251635  
; GENERAL INFORMATION:

APPLICANT: WAGNER, Fred W.  
APPLICANT: STOUT, Jay S.  
APPLICANT: HENRIKSEN, Dennis B.  
APPLICANT: PARTRIDGE, Bruce E.  
APPLICANT: HOLMQUIST, Bart  
APPLICANT: FRANK, Julie A.  
TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN  
FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN  
TITLE OF INVENTION: AND RELATED ANALOGS  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/139,819A  
FILING DATE: 25-AUG-1998  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/595,868  
FILING DATE: 06-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 089187/0144  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 867 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..864  
US-09-139-819A-11

Query Match 48.0%; Score 12; DB 4; Length 867;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttcgattcctcgt 16  
|||||

Db 532 TTCGATCCTCGT 543

RESULT 26  
US-08-726-725-5  
; Sequence 5, Application US/08726725  
; Patent No. 5773290  
; GENERAL INFORMATION:  
; APPLICANT: Gould, Michael N.  
; APPLICANT: Chen, Kai-Shun  
; TITLE OF INVENTION: MAMMARY GLAND-SPECIFIC PROMOTERS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 411 East Wisconsin Avenue  
; CITY: Milwaukee  
; STATE: Wisconsin  
; COUNTRY: U.S.A.

ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/726,725  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Baker, Jean C.  
REGISTRATION NUMBER: 35,433  
REFERENCE/DOCKET NUMBER: 960296,93863  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5709  
TELEFAX: (414) 271-3552  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 876 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-726-725-5

Query Match 48.0%; Score 12; DB 1; Length 876;  
Best local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atccctgcag 20  
|||||  
DB 310 ATCCTGTCAGG 321

RESULT 27  
US-09-126-109-5/c  
Sequence 5, Application US/09126109  
Patent No. 6171856  
GENERAL INFORMATION:  
APPLICANT: Thigpen, Anice  
APPLICANT: Hohmeier, Hans-Ewald  
APPLICANT: Newgard, Christopher B.  
APPLICANT: Unger, Roger H.  
APPLICANT: Shimabukuro, Michio  
APPLICANT: Chen, Guaxun  
APPLICANT: Rhodes, Christopher J.  
APPLICANT: Hugl, Sigfrun R.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING  
TO NO-MEDITATED CYTOTOXICITY  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/126,109  
FILING DATE: 30-JUL-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,092  
FILING DATE: 30-JUL-1997  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US Unknown  
FILING DATE: 03-MAR-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: McMillian, Nabeela R.  
REGISTRATION NUMBER: P-43,363  
REFERENCE/DOCKET NUMBER: UTSD:560  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 911 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-126-109-5

Query Match 48.0%; Score 12; DB 4; Length 911;  
Best local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtgcag 25  
|||||  
DB 512 CGTCAGGTGCAG 501

RESULT 28  
PCT-US93-06251-3/c  
Sequence 3, Application PC/TUS9306251  
GENERAL INFORMATION:  
APPLICANT: Wickstrom, Eric and Rife, Jason P.  
TITLE OF INVENTION: Trivalent synthesis of Oligonucleotides Containing  
Stereospecific Alkylphosphonates and Arylphosphonates  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: NY  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/06251  
FILING DATE: 19930630  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8586  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 911 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US93-06251-3

Query Match 48.0%; Score 12; DB 5; Length 911;  
Best local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



OY 14 cgtcagatgcag 25  
|||||  
Db 512 CCGTCAGTGCAG 501

## RESULT 29

US-07-952-817-10  
; Sequence 10, Application US/07952817  
; Patent No. 5356796

## GENERAL INFORMATION:

APPLICANT: Keller, John W.

TITLE OF INVENTION: A Repressor Protein and Gene for Regulating

EXPRESSION OF POLYPEPTIDES AND ITS USE IN THE PREPARATION OF

2,2-DIALKYLGLYCINE DECARBOXYLASE OF PSEUDOMONAS CEPACIA

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finegan, Henderson, Farbow, Garrett &

STREET: 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: D.C.

COUNTRY: US

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/952,817

FILING DATE: 19920928

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.

REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 01120.0002-01000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-408-4000

TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1411 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-07-952-817-10

Query Match 48.0%; Score 12; DB 1; Length 1411;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatctgcattcc 12  
|||||

Db 655 GCATTGCATCC 666

## RESULT 30

5210025-3

; Patent No. 5210025

APPLICANT: KELLER, JOHN W.

TITLE OF INVENTION: REPRESSOR PROTEIN GENE FOR REGULATING

EXPRESSION OF POLYPEPTIDES AND ITS USE IN THE PREPARATION OF

2,2-DIALKYLGLYCINE DECARBOXYLASE OF PSEUDOMONAS CEPACIA

NUMBER OF SEQUENCES: 18

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/501,814

FILING DATE: 30-MAR-1990

SEQ ID NO:3;

LENGTH: 1411

5210025-3

Query Match 48.0%; Score 12; DB 6; Length 1411;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatctgcattcc 12  
|||||

Db 655 gcatctgcattcc 666

## RESULT 31

US-09-248-528-3/c

; Sequence 3, Application US/09248528C

; Patent No. 6153415

GENERAL INFORMATION:

APPLICANT: Oriol, Patrick J

APPLICANT: Padmakumar, Rugmini

APPLICANT: Kim, Sang H

TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite

TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus

FILE REFERENCE: MSU 4.1-401

CURRENT APPLICATION NUMBER: US/09/248,528C

CURRENT FILING DATE: 1999-02-10

EARLIER APPLICATION NUMBER: 60/083,485

EARLIER FILING DATE: 1998-04-29

NUMBER OF SEQ ID NOS: 18

SOFTWARE: Patent Ver. 2.0

SEQ ID NO 3

LENGTH: 1516

TYPE: DNA

ORGANISM: Bacillus pallidus

FEATURE:

NAME/KEY: rRNA

LOCATION: (1)..(1516)

PUBLICATION INFORMATION:

DATABASE ACCESSION NUMBER: Z26930/GenBank

DATABASE ENTRY DATE: 1997-05-14

US-09-248-528-3

Query Match 48.0%; Score 12; DB 3; Length 1516;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcctcgtcaggt 21  
|||||

Db 484 TCCTCGTCAGGT 473

## RESULT 32

US-09-549-108-3/c

; Sequence 3, Application US/09549108

; Patent No. 6214603

GENERAL INFORMATION:

APPLICANT: Oriol, Patrick J

APPLICANT: Padmakumar, Rugmini

APPLICANT: Kim, Sang H

TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite

TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus

FILE REFERENCE: MSU 4.1-486

CURRENT APPLICATION NUMBER: US/09/549,108

CURRENT FILING DATE: 2000-04-13

PRIOR APPLICATION NUMBER: 60/083,485

PRIOR FILING DATE: 1998-04-29

PRIOR APPLICATION NUMBER: 09/248,528

PRIOR FILING DATE: 1999-02-10

NUMBER OF SEQ ID NOS: 18

SOFTWARE: Patent Ver. 2.0

SEQ ID NO 3

LENGTH: 1516

TYPE: DNA

ORGANISM: Bacillus pallidus

```
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1516)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Z26930/GenBank
DATABASE ENTRY DATE: 1997-05-14
US-09-549-108-3
```

```
Query Match      48.0%; Score 12; DB 4; Length 1516;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      10 tctctgcaggt 21
        |||
Db      484 TCCTGTCAGGT 473
```

```
RESULT 33
US-09-549-111-3/c
Sequence 3, Application US/09549111
Patent No. 6228633
GENERAL INFORMATION:
APPLICANT: Oriol, Patrick J
APPLICANT: Padmakumar, Rugmini
APPLICANT: Kim, Sang H
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
FILE REFERENCE: MSU 4.1-489
CURRENT APPLICATION NUMBER: US/09/549,111
CURRENT FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/083,485
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 09/248,528
PRIOR FILING DATE: 1999-02-10
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1516
TYPE: DNA
ORGANISM: Bacillus pallidus
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1516)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Z26930/GenBank
DATABASE ENTRY DATE: 1997-05-14
US-09-549-111-3
```

```
Query Match      48.0%; Score 12; DB 4; Length 1516;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      10 tctctgcaggt 21
        |||
Db      484 TCCTGTCAGGT 473
```

```
RESULT 34
US-09-549-106-3/c
Sequence 3, Application US/09549106
Patent No. 6242242
GENERAL INFORMATION:
APPLICANT: Oriol, Patrick J
APPLICANT: Padmakumar, Rugmini
APPLICANT: Kim, Sang H
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
FILE REFERENCE: MSU 4.1-487
CURRENT APPLICATION NUMBER: US/09/549,106
CURRENT FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/083,485
```

```
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 09/248,528
PRIOR FILING DATE: 1999-02-10
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1516
TYPE: DNA
ORGANISM: Bacillus pallidus
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1516)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Z26930/GenBank
DATABASE ENTRY DATE: 1997-05-14
US-09-549-106-3
```

```
Query Match      48.0%; Score 12; DB 4; Length 1516;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      10 tctctgcaggt 21
        |||
Db      484 TCCTGTCAGGT 473
```

```
RESULT 35
US-09-248-528-2/c
Sequence 2, Application US/09248528C
Patent No. 6153415
GENERAL INFORMATION:
APPLICANT: Oriol, Patrick J
APPLICANT: Padmakumar, Rugmini
APPLICANT: Kim, Sang H
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
FILE REFERENCE: MSU 4.1-401
CURRENT APPLICATION NUMBER: US/09/248,528C
CURRENT FILING DATE: 1999-02-10
EARLIER APPLICATION NUMBER: 60/083,485
EARLIER FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1517
TYPE: DNA
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1517)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Z26929
DATABASE ENTRY DATE: 1998-07-02
US-09-248-528-2
```

```
Query Match      48.0%; Score 12; DB 3; Length 1517;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      10 tctctgcaggt 21
        |||
Db      487 TCCTGTCAGGT 476
```

```
RESULT 36
US-09-549-108-2/c
Sequence 2, Application US/09549108
Patent No. 6214603
GENERAL INFORMATION:
APPLICANT: Oriol, Patrick J
APPLICANT: Padmakumar, Rugmini
```

```
APPLICANT: Kim, Sang H
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite
FILE REFERENCE: MSU 4.1-486
CURRENT FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/083,485
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 09/248,528
PRIOR FILING DATE: 1999-02-10
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1517
TYPE: DNA
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1517)
DATABASE ACCESSION NUMBER: Z26929
DATABASE ENTRY DATE: 1998-07-02
US-09-549-108-2
```

```
Query Match
Best Local Similarity 48.0%; Score 12; DB 4; Length 1517;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 tcctcgtcagct 21
      |||||
Db 487 TCCTCGTCAGCT 476
```

```
RESULT 37
US-09-549-111-2/c
Sequence 2, Application US/09549111
GENERAL INFORMATION:
APPLICANT: Oriel, Patrick J
APPLICANT: Padmakumar, Rugmini
APPLICANT: Kim, Sang H
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite
FILE REFERENCE: MSU 4.1-489
CURRENT FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/083,485
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 09/248,528
PRIOR FILING DATE: 1999-02-10
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1517
TYPE: DNA
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1517)
DATABASE ACCESSION NUMBER: Z26929
DATABASE ENTRY DATE: 1998-07-02
US-09-549-111-2
```

```
Query Match
Best Local Similarity 48.0%; Score 12; DB 4; Length 1517;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 tcctcgtcagct 21
      |||||
Db 487 TCCTCGTCAGCT 476
```

```
RESULT 38
US-09-549-106-2/c
Sequence 2, Application US/09549106
PATENT NO. 6242242
GENERAL INFORMATION:
APPLICANT: Oriel, Patrick J
APPLICANT: Padmakumar, Rugmini
APPLICANT: Kim, Sang H
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite
FILE REFERENCE: MSU 4.1-487
CURRENT FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/083,485
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 09/248,528
PRIOR FILING DATE: 1999-02-10
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1517
TYPE: DNA
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1517)
DATABASE ACCESSION NUMBER: Z26929
DATABASE ENTRY DATE: 1998-07-02
US-09-549-106-2
```

```
Query Match
Best Local Similarity 48.0%; Score 12; DB 4; Length 1517;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 tcctcgtcagct 21
      |||||
Db 487 TCCTCGTCAGCT 476
```

```
RESULT 39
US-08-987-151-4
Sequence 4, Application US/08987151
PATENT NO. 6162617
GENERAL INFORMATION:
APPLICANT: Jaskunas Jr., Stanley R.
APPLICANT: Zhao, Genshi
APPLICANT: Peery, Robert B.
APPLICANT: Burgelt, Stanley G.
APPLICANT: Rostock Jr., Paul R.
TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESS: Eli Lilly and Company
CITY: Indianapolis
STATE: Indiana
COUNTRY: U.S.
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/987,151
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Webster, Thomas D.
REGISTRATION NUMBER: 39,872
REFERENCE/DOCKET NUMBER: X-11748
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3334  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1602 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-987-151-4

Query Match 48.0%; Score 12; DB 4; Length 1602;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tcgtcaggtgca 24  
|||||  
Db 1398 TCGTCAGGTGCA 1409

RESULT 40  
US-08-987-151-1  
Sequence 1, Application US/08987151  
Patent No. 6162617  
GENERAL INFORMATION:  
APPLICANT: Jaskunas Jr., Stanley R.  
APPLICANT: Zhao, Genshl  
APPLICANT: Peery, Robert B.  
APPLICANT: Burgett, Stanley G.  
APPLICANT: Rostock Jr., Paul R.  
TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence  
NUMBER OF SEQUENCES: 4  
CURRENT APPLICATION DATA:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: U.S.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/987,151  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Webster, Thomas D.  
REGISTRATION NUMBER: 39,872  
REFERENCE/DOCKET NUMBER: X-11748  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3334  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1785 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1782  
US-08-987-151-1

Query Match 48.0%; Score 12; DB 4; Length 1785;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tcgtcaggtgca 24  
|||||  
Db 1305 TCGTCAGGTGCA 1316

RESULT 41  
US-08-987-151-3  
Sequence 3, Application US/08987151  
Patent No. 6162617  
GENERAL INFORMATION:  
APPLICANT: Jaskunas Jr., Stanley R.  
APPLICANT: Zhao, Genshl  
APPLICANT: Peery, Robert B.  
APPLICANT: Burgett, Stanley G.  
APPLICANT: Rostock Jr., Paul R.  
TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence  
NUMBER OF SEQUENCES: 4  
CURRENT APPLICATION DATA:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: U.S.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/987,151  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Webster, Thomas D.  
REGISTRATION NUMBER: 39,872  
REFERENCE/DOCKET NUMBER: X-11748  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3334  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1785 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-987-151-3

Query Match 48.0%; Score 12; DB 4; Length 1785;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 13 tcgtcaggtgca 24  
|||||  
Db 1305 UCGCAGUGUCA 1316

RESULT 42  
US-08-365-486A-16/c  
Sequence 16, Application US/08365486A  
Patent No. 5834306  
GENERAL INFORMATION:  
APPLICANT: Webster, Keith A.  
APPLICANT: Bishopric, Nanette H.  
TITLE OF INVENTION: Tissue Specific Hypoxia Regulated

TITLE OF INVENTION: Therapeutic Constructs  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/365,486A  
FILING DATE: 23-DEC-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 8255-0018  
TELEPHONE: (415) 324-0960  
TELEFAX: (415) 324-0880  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1846 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: bcl-2 fusion gene; Seto, et al.,  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 887..1606  
US-08-365-486A-16

Query Match 48.0%; Score 12; DB 2; Length 1846;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cgtcaggtgcag 25  
|||||  
Db 1252 CGTCAGGTGCAG 1241

RESULT 43  
US-08-880-342-16/C  
Sequence 16, Application US/08880342  
Patent No. 6218179  
GENERAL INFORMATION:  
APPLICANT: Webster, Keith A.  
APPLICANT: Bishopric, Nanette H.  
APPLICANT: Murphy, Brian  
APPLICANT: Laderoute, Keith R.  
APPLICANT: Green, Christopher J.  
TITLE OF INVENTION: Tissue Specific Hypoxia Regulated  
TITLE OF INVENTION: Therapeutic Constructs  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/880,342  
FILING DATE: 23-JUN-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/IB95/00996  
FILING DATE: 13-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/365,486  
FILING DATE: 23-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 8255-0018.30  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1846 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: bcl-2 fusion gene; Seto, et al.,  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 887..1606  
US-08-880-342-16

Query Match 48.0%; Score 12; DB 4; Length 1846;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cgtcaggtgcag 25  
|||||  
Db 1252 CGTCAGGTGCAG 1241

RESULT 44  
US-09-026-587-2  
Sequence 2, Application US/09026587  
Patent No. 5912128  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN ENA/VASP-LIKE PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSRO for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/026,587  
FILING DATE: Filed Herewith

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0471 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1889 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HEAONOT03  
CLONE: 3089412  
US-09-026-587-2

Query Match 48.0%; Score 12; DB 2; Length 1889;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 cgtcaggtgcag 25  
|||||  
Db 496 CGTCAGGTGCAG 507

RESULT 45  
US-09-227-420-2  
Sequence 2, Application US/09227420  
Patent No. 5990087  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN ENA/VASP-LIKE PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/227,420  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/026,587  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0471 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1889 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HEAONOT03  
CLONE: 3089412  
US-09-227-420-2

Query Match 48.0%; Score 12; DB 2; Length 1889;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 cgtcaggtgcag 25  
|||||  
Db 496 CGTCAGGTGCAG 507

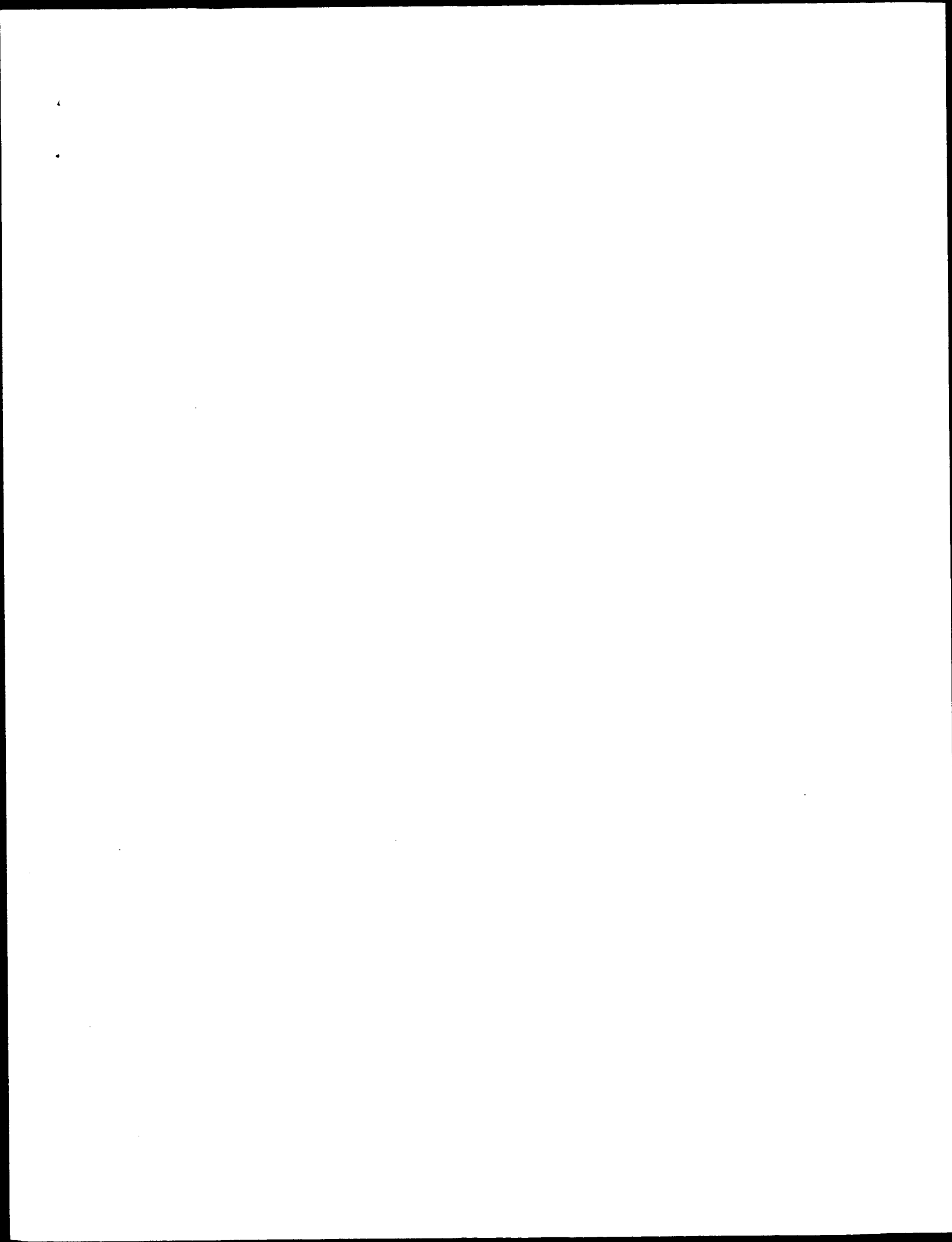
Search completed: October 9, 2001, 15:55:41  
Job time: 13217 sec

Wed Oct 10, 07:46:09 2001

us-09-396-196f-8.oli.rni

---

Page 19





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:34 ; Search time 5323.87 Seconds  
(without alignments)  
44.389 Million cell updates/sec

Title: US-09-396-196F-8

Perfect score: 25

Sequence: 1 gcattcgatccctgcgtgcag 25

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 322778

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

```
117: gb_est148:*
118: gb_est149:*
119: gb_est150:*
120: gb_est151:*
121: gb_est152:*
122: gb_est153:*
123: gb_est154:*
124: gb_est155:*
125: gb_est156:*
126: gb_est157:*
127: gb_est158:*
128: gb_est159:*
129: gb_est160:*
130: gb_est161:*
131: gb_est162:*
132: gb_est163:*
133: gb_est164:*
134: gb_est165:*
135: gb_est166:*
136: gb_est167:*
137: gb_est168:*
138: gb_est169:*
139: gb_est170:*
140: gb_est171:*
141: gb_est172:*
142: gb_est173:*
143: gb_est174:*
144: gb_est175:*
145: gb_est176:*
146: gb_est177:*
147: gb_est178:*
148: gb_est179:*
149: gb_est180:*
150: gb_est181:*
151: gb_est182:*
152: gb_est183:*
153: gb_est184:*
154: gb_est185:*
155: gb_est186:*
156: gb_est187:*
157: gb_est188:*
158: gb_est189:*
159: gb_est190:*
160: gb_est191:*
161: gb_est192:*
162: gb_est193:*
163: gb_est194:*
164: gb_est195:*
165: gb_est196:*
166: gb_est197:*
167: gb_est198:*
168: gb_est199:*
169: gb_est200:*
170: gb_est201:*
171: gb_est202:*
172: gb_est203:*
173: gb_est204:*
174: gb_est205:*
175: gb_est206:*
176: gb_est207:*
177: gb_est208:*
178: gb_est209:*
179: gb_est210:*
180: gb_est211:*
181: gb_est212:*
182: gb_est213:*
183: gb_est214:*
184: gb_est215:*
185: gb_est216:*
186: gb_est217:*
187: gb_est218:*
188: gb_est219:*
189: gb_est220:*

190: gb_est110:*
191: gb_est111:*
192: gb_hlc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 16    | 64.0        | 300    | AU113702  | AU113702 AU113702  |
| 2          | 16    | 64.0        | 300    | AU116641  | AU116641 AU116641  |
| 3          | 16    | 64.0        | 175    | BG269235  | BG269235 LO-3267T3 |
| 4          | 15    | 60.0        | 431    | AM925187  | AM925187 WSI-76-EO |
| 5          | 15    | 60.0        | 519    | BE442332  | BE442332 925017G02 |
| 6          | 15    | 60.0        | 542    | AZ726169  | AZ726169 RPCI-24-7 |
| 7          | 15    | 60.0        | 543    | AZ168021  | AZ168021 SP-0103-B |
| 8          | 15    | 60.0        | 658    | C98409    | C98409 C98409_Rice |
| 9          | 15    | 60.0        | 668    | AU165518  | AU165518 AU165518  |
| 10         | 15    | 60.0        | 713    | AU094990  | AU094990 AU094990  |
| 11         | 15    | 60.0        | 1096   | BF144032  | BF144032 601791288 |
| 12         | 15    | 60.0        | 1126   | BE795069  | BE795069 601592885 |
| 13         | 14    | 56.0        | 195    | W06020    | W06020 T9EST787F1  |
| 14         | 14    | 56.0        | 196    | AI210255  | AI210255 h4a02a1.r |
| 15         | 14    | 56.0        | 208    | BF713183  | BF713183 MI-P-01-a |
| 16         | 14    | 56.0        | 257    | BG408765  | BG408765 gb76c07.Y |
| 17         | 14    | 56.0        | 266    | AV537540  | AV537540 AV537540  |
| 18         | 14    | 56.0        | 288    | BG382504  | BG382504 298426 MA |
| 19         | 14    | 56.0        | 291    | BF994483  | BF994483 CM3-GN010 |
| 20         | 14    | 56.0        | 304    | BF402559  | BF402559 UI-R-CAO- |
| 21         | 14    | 56.0        | 305    | C72075    | C72075 C72075_Rice |
| 22         | 14    | 56.0        | 327    | AI441334  | AI441334 sa55d11.Y |
| 23         | 14    | 56.0        | 360    | C39749    | C39749 C39749_Yu1  |
| 24         | 14    | 56.0        | 368    | AM699577  | AM699577 gb08f06.Y |
| 25         | 14    | 56.0        | 378    | AL370675  | AL370675 MEB439C04 |
| 26         | 14    | 56.0        | 381    | FR0035439 | AL122961 Fugu rubr |
| 27         | 14    | 56.0        | 420    | BE021456  | BE021456 sm49c06.Y |
| 28         | 14    | 56.0        | 421    | AI938935  | AI938935 sc63b07.Y |
| 29         | 14    | 56.0        | 429    | A0205063  | A0205063 HS-3226-B |
| 30         | 14    | 56.0        | 430    | AI272538  | AI272538 uk05h06.Y |
| 31         | 14    | 56.0        | 432    | AQ129442  | AQ129442 HS-3045-A |
| 32         | 14    | 56.0        | 454    | AK017623  | AK017623 Mus muscu |
| 33         | 14    | 56.0        | 473    | AQ815087  | AQ815087 HS-5261-B |
| 34         | 14    | 56.0        | 486    | AU066422  | AU066422 AU066422  |
| 35         | 14    | 56.0        | 494    | AI507864  | AI507864 sa88b04.Y |
| 36         | 14    | 56.0        | 518    | BF007983  | BF007983 1619492.Y |
| 37         | 14    | 56.0        | 521    | AL379335  | AL379335 MEB44F11  |
| 38         | 14    | 56.0        | 524    | BF634261  | BF634261 NC084C10D |
| 39         | 14    | 56.0        | 532    | CNS04RAN  | AL303656 Tetradon  |
| 40         | 14    | 56.0        | 548    | AL381376  | AL381376 MEB019F0  |
| 41         | 14    | 56.0        | 559    | AQ528128  | AQ528128 RPCI-11-3 |
| 42         | 14    | 56.0        | 586    | BG080788  | BG080788 H3056F11- |
| 43         | 14    | 56.0        | 610    | AQ655910  | AQ655910 Sheared D |
| 44         | 14    | 56.0        | 642    | AM953979  | AM953979 EST365944 |
| 45         | 14    | 56.0        | 644    | BG067650  | BG067650 H3056F11- |

## ALIGNMENTS

| RESULT     | 1  | LOCUS   | AU113702       | 300 bp | mRNA | EST | 19-OCT-2000 |
|------------|--|---|----------------|--------|------|-----|-------------|
| DEFINITION | AU113702   | unpublished oligo-capped cDNA library         | Caenorhabditis |        |      |     |             |
| ACCESSION  | AU113702   | elegans cDNA clone yk710f7 3', mRNA sequence. |                |        |      |     |             |
| VERSION    | AU113702.1   | GI:10927269                                   |                |        |      |     |             |
| KEYWORDS   | EST.   |   |                |        |      |     |             |
| SOURCE     | Caenorhabditis elegans.  |   |                |        |      |     |             |
| ORGANISM   | Caenorhabditis elegans.  |   |                |        |      |     |             |
| REFERENCE  | 1  | (bases 1 to 300)                              |                |        |      |     |             |
| AUTHORS    | Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y. |   |                |        |      |     |             |
| TITLE      | A complementary view of the C. elegans genome                          |   |                |        |      |     |             |
| JOURNAL    | Unpublished (2000)   |   |                |        |      |     |             |

## COMMENT

Contact: Yuji Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.

## FEATURES

Location/Qualifiers  
1..300

/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone="yk710f7"  
/clone\_lib="unpublished oligo-capped cDNA library"  
/sex="hermaphrodite"  
/tissue\_type="whole animal"  
/dev\_stage="varied"

## BASE COUNT

97 a 50 c 60 g 93 t

## Query Match

Best Local Similarity 64.0%; Score 16; DB 107; Length 300;  
Pred. No. 6.3; Mismatches 0; Indels 0; Gaps 0;

## Matches

16; Conservative

## Db

211 TTTCGATCCTCGTCAG 226

## RESULT

2

## LOCUS

AU116641

## DEFINITION

AU116641 unpublished oligo-capped cDNA library

## ACCESSION

AU116641

## VERSION

AU116641.1 GI:10930208

## KEYWORDS

Caenorhabditis elegans.

## SOURCE

Caenorhabditis elegans.

## ORGANISM

Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea

## REFERENCE

1 (bases 1 to 300)  
; Rhabdilitida; Pelodierinae; Caenorhabditis.

## AUTHORS

Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.

## TITLE

A complementary view of the C. elegans genome

## JOURNAL

Unpublished (2000)

## COMMENT

Gene Library Lab  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.

## FEATURES

Location/Qualifiers  
1..300

## source

/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone="yk710f7"  
/clone\_lib="unpublished oligo-capped cDNA library"  
/sex="hermaphrodite"  
/tissue\_type="whole animal"  
/dev\_stage="varied"

## BASE COUNT

97 a 51 c 58 g 92 t 2 others

## ORIGIN

1

## Query Match

Best Local Similarity 64.0%; Score 16; DB 107; Length 300;  
Pred. No. 6.3; Mismatches 0; Indels 0; Gaps 0;

## Matches

16; Conservative

## Db

4 TTTCGATCCTCGTCAG 19

Db 221 TTTCGATCCTGTCAGT 236

RESULT 3  
LOCUS BG269235 695 bp mRNA EST 20-FEB-2001  
DEFINITION L0-3267T3 Ice plant Lambda Uni-Zap XR expression library, 0 hours  
NACl treatment Mesembryanthemum crystallinum cDNA clone L0-3267 5',  
mRNA sequence.

ACCESSION BG269235  
VERSION BG269235.1 GI:12975075  
KEYWORDS EST.  
SOURCE common ice plant.  
ORGANISM Mesembryanthemum crystallinum

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
Spermatophyta; Magnoliophyta: eudicotyledons: core eudicots;  
Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.  
1 (bases 1 to 695)  
AUTHORS Cushman, J.C.  
TITLE An expressed sequence tag database for the common ice plant,  
Mesembryanthemum crystallinum  
JOURNAL Unpublished (1997)  
COMMENT Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu  
PCR PRIMERS  
FORWARD: T7  
BACKWARD: T3  
Plate: L0-33 row: F column: 7  
Seq primer: T3  
High quality sequence stop: 350  
POLYA-No.

FEATURES  
source 1..695 Location/Qualifiers  
/organism="Mesembryanthemum crystallinum"  
/db\_xref="taxon:3544"  
/clone\_lib="L0-3267"  
/clone\_lib="Ice plant Lambda Uni-Zap XR expression library  
/0 hours NACl treatment"  
/tissue\_type="leaf"  
/dev\_stage="Six week old"  
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site\_1:  
ECORI; Site\_2: XhoI"  
BASE COUNT 221 a 148 c 158 g 168 t  
ORIGIN

Query Match 64.0%; Score 16; DB 175; Length 695;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 tcgattcctgcaggt 21  
Db 104 TCGATCCTGTCAGT 119

RESULT 4  
LOCUS AW925187/c 431 bp mRNA EST 19-JUL-2000  
DEFINITION WSI\_76\_E02.bl\_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,  
mRNA sequence.  
ACCESSION AW925187  
VERSION AW925187.1 GI:8091013  
KEYWORDS EST.  
SOURCE sorghum.  
ORGANISM Sorghum bicolor  
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

REFERENCE 1 (bases 1 to 431)  
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt,  
L.H.  
TITLE An EST database from Sorghum: water-stressed plants  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude Polya, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.  
Seq primer: JEN REV  
High quality sequence stop: 339  
POLYA-No.

FEATURES  
source 1..431 Location/Qualifiers  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_lib="Water-stressed 1 (WS1)"  
/note="Organ: Mix of 5-week old plants on days 7 & 8 after  
water was withheld; Vector: Lambda Zap; Site\_1: XhoI;  
Site\_2: EcoRI; The library was made from polyA RNA in the  
cloning vector Lambda Zap II. Clones to be sequenced were  
prepared by mass excision."  
BASE COUNT 98 a 135 c 106 g 92 t  
ORIGIN

Query Match 60.0%; Score 15; DB 122; Length 431;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 gattcctgcaggtg 22  
Db 419 GATCCTGTCAGTGTG 405

RESULT 5  
LOCUS BE442332 519 bp mRNA EST 25-JUL-2000  
DEFINITION 925017602.xl C. reinhardtii CC-2290, normalized, Lambda Zap II  
Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION BE442332  
VERSION BE442332.1 GI:9441847  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii.  
Eukaryota: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonadales.  
1 (bases 1 to 519)  
Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,  
Mcdermott, J.P., Silflow, C., Stern, D. and Surzycki, R.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants; project phase 2  
Unpublished (2000)  
JOURNAL Contact: Elizabeth H. Harris  
COMMENT DCM Box 91000  
Duke University  
Durham, NC 27708-1000, USA  
Tel: 919 613 8164  
Fax: 919 613 8177  
Email: chlamy@duke.edu.

FEATURES  
source 1..519 Location/Qualifiers  
/organism="Chlamydomonas reinhardtii"  
/strain="CC-2290 wild type mt- S1 D2"  
/db\_xref="taxon:3055"

/clone\_lib="C. reinhardtii CC-2290, normalized, lambda zap  
11"

/note="Vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2:  
XhoI; This library was constructed by John Davies and  
Jeffrey McDermott. RNA was isolated from strain CC-2290  
(Minnesota isolate of *C. reinhardtii*) grown to mid-log  
phase in TAP (acetate containing) medium in the light.  
POLYA mRNA was purified, and cDNA was synthesized and  
directionally cloned into lambda ZAP II (Stratagene) in  
the EcoRI (5') and XhoI (3') sites. pBluescript II SK-  
plasmids were excised from the lambda ZAP clones by  
superinfection with ExAssist (Stratagene) phage. The  
library was normalized using method 4 described in Ronaldo  
et al (1996) Genome Research 6: 791-806."

BASE COUNT 137 a 137 c 121 g 124 t

ORIGIN

Query Match 60.0%; Score 15; DB 167; Length 519;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctctgcagtgca 24  
|||||

Db 72 TCCTCGTCAGTGCA 86

RESULT 6

LOCUS A2726169 542 bp DNA GSS 24-JAN-2001

DEFINITION RPI-24-76J3.TV RPI-24 Mus musculus genomic clone RPI-24-76J3,  
DNA sequence.

ACCESSION A2726169

VERSION A2726169.1 GI:12473564

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 542)

AUTHORS Zhao, S., Niernan, W., Malek, J., Shatsman, S., Akınret, B., Levins, M.,  
Tsegaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregorgis, E.,  
Russell, D., de Jong, P. and Fraser, C. M.  
Mouse BAC End Sequences from Library RPI-24  
Unpublished (1999)

OTHER GSS: RPI-24-76J3.TJ

COMMENT Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPI-24. For BAC  
library availability, please contact Pieter de Jong  
(pjejong@tigr.org). Clones may be purchased from BACPAC  
Resources (<http://www.chori.org/bacpac/orderingframe.html>). BAC end  
Plate: 76 row: J column: 3  
Seq primer: T7  
Class: BAC ends.

FEATURES

Source 1..542

Location/Qualifiers

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPI-24-76J3"

/clone\_lib="RPI-24"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="Vector: pPARBAC1; Site\_1: BamHI; Site\_2: BamHI.  
RPI-24 Mouse BAC Library produced by Pieter de Jong. The  
library was cloned in the pPARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J  
DNA."

BASE COUNT 152 a 126 c 132 g 132 t

ORIGIN

Query Match 60.0%; Score 15; DB 248; Length 542;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcattcctgcag 20  
|||||

Db 149 TCATCTCTCTGTCAGG 163

RESULT 7

LOCUS A2168021 543 bp DNA GSS 29-AUG-2000

DEFINITION SP\_0103\_B1\_F05\_T7A Strongylocentrotus purpuratus, purple sea urchin  
, sperm genomic BAC library Strongylocentrotus purpuratus genomic  
clone Plate-103 Col-9 Row=L, DNA sequence.

ACCESSION A2168021

VERSION A2168021.1 GI:8338389

KEYWORDS GSS.

SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus

REFERENCE Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
Echinoidea; Euechinoidea; Echinoidea; Echinoidea;  
Strongylocentrotidae; Strongylocentrotus.  
1 (bases 1 to 543)

AUTHORS Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,  
Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray,  
G.A., Ettensohn, C.A., Lohrman, H., Britten, R.J., Davidson, E.H. and  
Hood, L.  
A sea urchin genome project: Sequence scan, virtual map, and  
additional resources  
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

CONTACT: Cameron, RA, Davidson, EH, Hood, L  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu

Plate: 103 row: L column: 9  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 543.

FEATURES

Source 1..543

Location/Qualifiers

/organism="Strongylocentrotus purpuratus"

/db\_xref="taxon:7668"

/clone="Plate=103 Col=9 Row=L"

/clone\_lib="Strongylocentrotus purpuratus, purple sea  
urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli  
DH10B"

BASE COUNT 149 a 116 c 116 g 157 t 5 others

ORIGIN

Query Match 60.0%; Score 15; DB 239; Length 543;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atctctgcagtgca 23  
|||||

Db 115 ATCTCTCTGTCAGTGC 101

RESULT 8

C98409

LOCUS C98409 658 bp mRNA EST 19-OCT-1998  
 DEFINITION C98409 Rice panicle at flowering stage Oryza sativa cDNA clone  
 accession E0106\_62, mRNA sequence.  
 VERSION C98409.1 GI:3761161  
 KEYWORDS EST.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 658)  
 AUTHORS Sasaki,T. and Yamamoto,K.  
 TITLE Rice cDNA from panicle at flowering stage  
 JOURNAL Unpublished (1996)  
 COMMENT contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program  
 2-1-2 Kannondai, Tsukuba  
 Ibaraki,  
 Japan 305  
 Tel: 0298-38-7441  
 Fax: 0298-38-7468  
 Email: tsasaki@abr.affrc.go.jp  
 PROJECT = "RGP".

FEATURES  
 source  
 1. .658  
 /organism="Oryza sativa"  
 /strain="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="E0106\_62"  
 /clone\_id="Rice panicle at flowering stage"  
 /dev\_stage="flowering stage"  
 /note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 161 a 159 c 155 g 179 t 4 others  
 ORIGIN

Query Match 60.0%; Score 15; DB 156; Length 658;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcctcgtaagtcca 24  
 |||||||||||||||  
 DB 177 TCCTCGTCAGGTGCA 191

RESULT 9  
 LOCUS AU165518 668 bp mRNA EST 06-DEC-2000  
 DEFINITION AU165518 Rice panicle at flowering stage Oryza sativa cDNA clone  
 E2045, mRNA sequence.  
 VERSION AU165518  
 KEYWORDS EST.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 668)  
 AUTHORS Sasaki,T. and Yamamoto,K.  
 TITLE Rice cDNA from panicle at flowering stage (2000)  
 JOURNAL Unpublished (2000)  
 COMMENT contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program  
 2-1-2 Kannondai, Tsukuba  
 Ibaraki,  
 Japan 305  
 Tel: 0298-38-7441  
 Fax: 0298-38-7468  
 Email: tsasaki@abr.affrc.go.jp  
 PROJECT = "RGP".

FEATURES  
 source  
 1. .668  
 /organism="Oryza sativa"  
 /strain="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="E2045"  
 /clone\_id="Rice panicle at flowering stage"  
 /dev\_stage="flowering stage"  
 /note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 160 a 162 c 162 g 179 t 5 others  
 ORIGIN

Query Match 60.0%; Score 15; DB 108; Length 668;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcctcgtaagtcca 24  
 |||||||||||||||  
 DB 198 TCCTCGTCAGGTGCA 212

RESULT 10  
 LOCUS AU094990 713 bp mRNA EST 30-JUN-2000  
 DEFINITION AU094990 Rice panicle (between 3cm to 10cm) Oryza sativa cDNA clone  
 E41284, mRNA sequence.  
 VERSION AU094990.1 GI:8857672  
 KEYWORDS EST.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 713)  
 AUTHORS Sasaki,T. and Yamamoto,K.  
 TITLE Rice cDNA from panicle (between 3cm to 10cm) (2000)  
 JOURNAL Unpublished (2000)  
 COMMENT contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program  
 2-1-2 Kannondai, Tsukuba  
 Ibaraki,  
 Japan 305  
 Tel: 0298-38-7441  
 Fax: 0298-38-7468  
 Email: tsasaki@abr.affrc.go.jp  
 PROJECT = "RGP".  
 E41284\_82.

FEATURES  
 source  
 1. .713  
 /organism="Oryza sativa"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="E41284"  
 /clone\_id="Rice panicle (between 3cm to 10cm)"  
 /clone\_type="panicle"  
 /tissue\_type="panicle"  
 /dev\_stage="panicle (between 3cm to 10cm)"

BASE COUNT 167 a 188 c 180 g 175 t 3 others  
 ORIGIN

Query Match 60.0%; Score 15; DB 107; Length 713;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcctcgtaagtcca 24  
 |||||||||||||||

Db 288 TCCTCTCAGCTCA 302

RESULT 11  
LOCUS BF144032/c  
DEFINITION 601791288F1 NCI\_CGAP\_Lu30 Mus musculus cDNA clone IMAGE:4022252 5', mRNA sequence.  
ACCESSION BF144032  
VERSION BF144032.1 GI:10983072  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1096)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LHAM9278 row: 0 column: 21  
High quality sequence stop: 630.  
Location/Qualifiers

FEATURES  
source 1..1096  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4022252"  
/clone\_1lb="NCI\_CGAP\_Lu30"  
/tissue\_type="tumor, metastatic to mammary"  
/lab\_host="DH10B"  
/note="Organ: Lung; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; transgenic model WNT-1, expression driven by MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 336 a 271 c 271 g 218 t

ORIGIN

Query Match 60.0%; Score 15; DB 145; Length 1096;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatctgcagctcg 15  
|||||

Db 1076 GCATTCGATCTCG 1062

RESULT 12  
LOCUS BE795069/c  
DEFINITION 601592885F1 NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:3946869 5', mRNA sequence.  
ACCESSION BE795069  
VERSION BE795069.1 GI:10216267  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1126)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LHCW807 row: j column: 22  
High quality sequence start: 10  
High quality sequence stop: 702.  
Location/Qualifiers

## FEATURES

source 1..1126  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3946869"  
/clone\_1lb="NIH\_MGC\_7"  
/tissue\_type="small cell carcinoma"  
/cell\_line="MGC3"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Lung; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(3). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 262 a 320 c 289 g 255 t

ORIGIN

Query Match 60.0%; Score 15; DB 140; Length 1126;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatctcgcagctg 22  
|||||

Db 853 GATCTCGTCAGCTG 839

RESULT 13  
LOCUS W06020/c  
DEFINITION TGSTZy87f10\_r1 TgRH tachyzoite cDNA Toxoplasma gondii cDNA clone tgyz87f10\_r1 5', mRNA sequence.  
ACCESSION W06020  
VERSION W06020.1 GI:1278733  
KEYWORDS EST.  
SOURCE Toxoplasma gondii.  
ORGANISM Toxoplasma gondii  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae; Toxoplasma.  
REFERENCE 1 (bases 1 to 195)  
AUTHORS Hehl, A., Manger, I., Marra, M., Sibley, L.D., Ajioke, J.A., Aslett, M.A., Dietrich, N., Dubuque, T., Hillier, L., Kucaba, T., Wan, K.L., Waterston, R.H. and Boothroyd, J.  
WashU-Merck-Stanford-NIH Toxoplasma EST project  
Unpublished (1996)  
CONTACT: Marra M  
WashU-Merck EST Project  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [toxowatson.wustl.edu](mailto:toxowatson.wustl.edu)  
David Sibley at toxowestborcim.wustl.edu for further information relating to organism, clone or library availability.  
Seq primer: T3  
High quality sequence stop: 148.  
Location/Qualifiers

## FEATURES

source 1..195  
/organism="Toxoplasma gondii"

/strain="RH"  
/db\_xref="taxon:5811"  
/clone="tgzy8f10.t1"  
/clone\_lib="TgRH Tachyzoite cDNA"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: lambda ZAP; Site\_1: EcoRI; Site\_2: XhoI;  
Toxoplasma RH strain tachyzoites were grown in human  
foreskin fibroblast cultures in vitro. The library was  
constructed by K.L. Wan, Cambridge University. cDNAs were  
synthesized from polyA RNAs by oligo d(T) priming and  
directionally cloned into the EcoRI to XhoI sites of the  
lambda ZAPII vector using the ZAP-cDNA synthesis kit  
(Stratagene) WARNING: the library contains a small  
percentage of cDNAs derived from the human host cells."

BASE COUNT 59 a 45 c 62 g 27 t 2 others

ORIGIN

Query Match 56.0%; Score 14; DB 189; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgcaccc 14  
|||||  
Db 181 GCATTCGATCCTC 168

RESULT 14  
AI210255 196 bp mRNA EST 19-OCT-1998  
LOCUS h4a02a1.r1 Aspergillus nidulans 24hr asexual developmental and  
DEFINITION vegetative cDNA lambda zap library Aspergillus nidulans cDNA clone  
h4a02a1 5', mRNA sequence.  
AI210255  
AI210255.1 GI:3772197

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
Aspergillus nidulans.  
Aspergillus nidulans  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; Emericella.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
1 (bases 1 to 196)  
Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,  
Prade, R. and Roe, B.  
An Aspergillus nidulans EST Database  
Unpublished (1998)  
Other ESTs: h4a02a1.f1  
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu  
Department of Chemistry and Biochemistry  
Advanced Center for Genome Technology, University of Oklahoma  
620 Parrington Oval, Norman, OK 73019, USA  
Tel: 405 325 4912  
Fax: 405 325 7762  
Email: broe@ou.edu  
We anticipate the future release of the cDNA clones to the Fungal  
Genetics Stock Center  
Seq primer: T3  
High quality sequence stop: 189.

FEATURES  
source  
Location/Qualifiers  
1..196  
/organism="Aspergillus nidulans"  
/strain="FGSC A26"  
/db\_xref="taxon:5072"  
/clone="h4a02a1"  
/clone\_lib="Aspergillus nidulans 24hr asexual  
developmental and vegetative cDNA lambda zap library"  
/tissue="vegetative mycelia, asexual structures"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript  
3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 43 a 60 c 37 g 55 t 1 others

ORIGIN

Query Match 56.0%; Score 14; DB 17; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gattccgtcagct 21  
|||||  
Db 32 GATTCGTGAGCT 45

RESULT 15  
BF713183 208 bp mRNA EST 02-JAN-2001  
LOCUS MI-P-01-adaq-d-09-1-UM.s1 MI-P-01 Sus scrofa cDNA clone  
DEFINITION MI-P-01-adaq-d-09-1-UM 3', mRNA sequence.  
BF713183  
BF713183.1 GI:12012658

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
Sus. scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
1 (bases 1 to 208)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
Contact: Tugale CK  
97044477  
Molecular Genetics Laboratory, Department of Animal Science  
Iowa State University  
201 Kildee Hall, Ames, IA 50011-3150, USA  
Tel: 5152844252  
Fax: 5152942401  
Email: cktugle@iastate.edu

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
non-normalized ovary at estrus day 0 library cDNA library  
Preparation: RJ Woods, JA Green, RS Prather 5142 Animal Science  
Research Center, Department of Animal Science, University of  
Missouri-Columbia, 65211 Clone distribution: clones will be  
available through Research Genetics (www.resgen.com) The following  
repetitive elements were found in this cDNA sequence: 145-208,  
>g12087|emb|X64127.1|SSPR S.scrofa DNA for SINE sequence SSPRE  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES  
source  
Location/Qualifiers  
1..208  
/organism="Sus scrofa"  
/strain="crossbred"  
/db\_xref="taxon:9823"  
/clone="MI-P-01-adaq-d-09-1-UM"  
/clone\_lib="MI-P-01"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3D-pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: EcoRI; The MI-P-01  
library is derived from ovary at estrus day 0. For a  
detailed description of the library from which this clone  
was derived, please visit our web site at  
http://pigst.genome.iastate.edu/  
TAG\_LIB=MI-P-01  
TAG\_TISSUE=ovary at estrus day 0  
TAG\_SEQ=GACGTA"

BASE COUNT 68 a 31 c 43 g 66 t

ORIGIN

Query Match 56.0%; Score 14; DB 168; Length 208;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Oy 8 gatccctcagct 21  
 |||||||  
 Db 189 GATCCTCCTCAGCT 176

RESULT 16  
 BG408765 257 bp mRNA EST 13-MAR-2001  
 LOCUS gb76c07.y1 Moss EST library PPG Physcomitrella patens cDNA clone  
 DEFINITION pep\_SOURCE\_ID=PPC\_CopyA-90613 5', mRNA sequence.

ACCESSION BG408765  
 VERSION BG408765.1 GI:13315054  
 KEYWORDS EST.

## ORGANISM

Physcomitrella patens.  
 Physcomitrella patens.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
 Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.

1 (bases 1 to 257)

Quatrano, R., Bashlades, S., Cove, D., Cuning, A., Knight, C., Clifton  
 S., Maria, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood  
 Steptoe, M., Gibbons, M., Harvey, N., Rilter, E., Jackson, Y., McCann, R.,  
 Leeds/Wash U Moss EST Project  
 Unpublished (1999)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Contact: Ralph Quatrano  
 Leeds/Wash U Moss EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu

Libraries were constructed by Dr. Stavros Bashlades as part of the  
 Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and  
 Washington Univ. in St. Louis (USA) DNA sequencing by: Washington  
 University Genome Sequencing Center for information on obtaining a  
 clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)  
 Seq primer: -40RP from Gibco  
 High quality: 40RP from Gibco

Location/Qualifiers

## FEATURES

## Source

1. 257  
 /organism="Physcomitrella patens"  
 /db\_xref="taxon:3218"  
 /clone="PEP\_SOURCE\_ID:PPG\_CopyA-90613"  
 /clone\_lib="Moss EST library PPG"  
 /tissue\_type="gametophore: 30 day old tissue,  
 ammonium-grown"  
 /lab\_host="DH10B"  
 /note="Vector: PAMPI; Construction of the cDNA library was  
 performed by Dr. W. Gregg Clark using a modification of  
 the cDNA synthesis protocol developed in the laboratory of  
 Dr. Michael Lovett by Dr. Yulia Korsunova (personal  
 communication). First polyA + RNA was isolated from total  
 gametophore RNA using oligo dT magnetic beads. Following  
 this, first strand cDNA synthesis was performed on the  
 bead-bound polyA + RNA, during which an oligonucleotide  
 anchor sequence was incorporated onto the 5'-ends of the  
 cDNA. PCR amplification was then used to synthesize the  
 second strand, to amplify the double stranded DNA, and to  
 incorporate dUTP containing sequences into the ends of the  
 double stranded cDNA. This DNA was size selected and  
 cloned into PAMPI using the CloneAMP PAMPI System (Life  
 Technologies, GibcoBRL) for cloning amplification products  
 by a non-restriction site dependant process. The cloning  
 was directional based on sequence asymmetry introduced at  
 the ends during PCR amplification. The 3' cDNA ends are  
 proximal to the NotI site of the multiple cloning site in  
 PAMPI. This annealing mixture was transformed into  
 chemically competent DH10B cells and selected for  
 ampicillin resistant growth. The resulting clones (about  
 330,000) were pooled to make the library."

## BASE COUNT

52 a 50 c 76 g 79 t

## ORIGIN

Query Match 56.0%; Score 14; DB 153; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 gatccctcagct 21  
 |||||||  
 Db 201 GATCCTCCTCAGCT 188

RESULT 17  
 AV537540 266 bp mRNA EST 06-SEP-2000  
 LOCUS AV537540 Arabidopsis thaliana roots Columbia Arabidopsis thaliana  
 DEFINITION cDNA clone R208C08F 3', mRNA sequence.

ACCESSION AV537540  
 VERSION AV537540.1 GI:8697823  
 KEYWORDS EST.

## ORGANISM

Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 266)

A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries  
 DNA Res. 7, 175-180 (2000)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

## FEATURES

## Source

1. 266  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="R208C08F"  
 /clone\_lib="Arabidopsis thaliana roots Columbia"  
 /tissue\_type="roots"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"

BASE COUNT 83 a 68 c 66 g 49 t

## ORIGIN

Query Match 56.0%; Score 14; DB 31; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 tttagcctcagct 17  
 |||||||  
 Db 157 TTTCGATCCTCAGCT 144

RESULT 18  
 BG382504 288 bp mRNA EST 12-MAR-2001  
 LOCUS BG382504  
 DEFINITION 288426 MARC 1P6G Sus scrofa cDNA 5', mRNA sequence.

ACCESSION BG382504  
 VERSION BG382504.1 GI:13306976  
 KEYWORDS EST.

## ORGANISM

## SOURCE

## pig.

## Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
 1 (bases 1 to 288)  
 Fahnenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,  
 Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W.

TITLE and Keele, J.W.  
Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine  
JOURNAL Unpublished (2000)  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt. trimmed with phred v0.980904e. Vector identified by cross\_match with the -mnscore 18 and -mismatch 12 options.

PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTCAGCAGC  
Plate: 4 row: 6 column: 9  
Seq primer: ATTAGCTGACACTATAG.

FEATURES  
Location/Qualifiers  
1..288

BASE COUNT  
ORIGIN  
126 a 57 c 56 g 49 t  
Query Match  
Best Local Similarity 100.0%; Score 14; DB 153; Length 288;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatccctgcaggt 21  
|||||  
Db 87 GATCCTGCTCAGGT 100

RESULT 19  
BF994483 291 bp mRNA EST 23-JAN-2001  
LOCUS BF994483  
DEFINITION CM3-GN0102-031100-459-h10 GN0102 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF994483  
VERSION BF994483.1 GI:12400806  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
1 (bases 1 to 291)  
Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Brines, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldmann, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&ct2=CM3-GN0102-031100-459-h10&t3=2000-11-03&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 177.  
Location/Qualifiers  
1..291

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0102"  
/dev\_stage="Adult"  
/note="Organ: placenta, normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT  
ORIGIN  
66 a 80 c 68 g 77 t

Query Match  
Best Local Similarity 100.0%; Score 14; DB 172; Length 291;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatccctgcaggt 21  
|||||  
Db 17 GATCCTGCTCAGGT 30

RESULT 20  
BF402559/c 304 bp mRNA EST 28-NOV-2000  
LOCUS BF402559/c  
DEFINITION UI-R-CA0-bhr-9-10-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone  
ACCESSION UI-R-CA0-bhr-9-10-0-UI 3', mRNA sequence.  
VERSION BF402559  
KEYWORDS BF402559.1 GI:11390534  
SOURCE EST.  
ORGANISM Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE  
AUTHORS  
1 (bases 1 to 304)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 9704477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@iuii.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLY-A=yes.

FEATURES  
source  
1..304  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone\_lib="UI-R-CA0-bhr-9-10-0-UI"  
/clone\_lib="UI-R-CA0"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CA0

library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

BASE COUNT 75 a 68 c 73 g 88 t  
TAG-SEO=None found\*

Query Match 56.0%; Score 14; DB 148; Length 304;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tectgcagtgagc 23  
|||||  
DB 246 TCCTCCTCAGCTGC 233

RESULT 21  
LOCUS C72075 305 bp mRNA EST 22-SEP-1997  
DEFINITION C72075 Rice panicle at flowering stage Oryza sativa cDNA clone  
E0927\_1A, mRNA sequence.  
ACCESSION C72075  
VERSION C72075.1 GI:2427612  
KEYWORDS EST.  
SOURCE Oryza sativa.

ORGANISM Oryza sativa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
REFERENCE 1 (bases 1 to 305)  
Sasaki, T. and Yamamoto, K.  
Rice cDNA from panicle at flowering stage  
Unpublished (1996)

AUTHORS National Institute of Agrobiological Resources  
TITLE Rice cDNA from panicle at flowering stage  
JOURNAL Unpublished (1996)  
COMMENT Contact: Takuji Sasaki  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: [tsasaki@agr.affrc.go.jp](mailto:tsasaki@agr.affrc.go.jp).

FEATURES  
source Location/Qualifiers  
1..305  
/organism="Oryza sativa"  
/strain="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="E0927\_1A"  
/clone\_1lb="Rice panicle at flowering stage"  
/dev\_stage="flowering stage"  
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 30 a 127 c 80 g 57 t 11 others  
ORIGIN

Query Match 56.0%; Score 14; DB 156; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 cgatctctgcagc 20  
|||||  
DB 192 CGATCTCTGCAGC 179

RESULT 22  
A1441334

LOCUS A1441334 327 bp mRNA EST 18-APR-2000  
DEFINITION sa55d11.y1 Gm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: O-METHYLTRANSFERASE ;, mRNA sequence.  
ACCESSION A1441334  
VERSION A1441334.1 GI:4288448  
KEYWORDS EST.  
SOURCE soybean.  
Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 327)  
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Pearson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.  
Public Soybean EST Project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: [clones@genomesystems.com](mailto:clones@genomesystems.com) or [info.genomesystems.com](http://info.genomesystems.com) web site: [www.genomesystems.com](http://www.genomesystems.com)  
Insert Length: 1286 Std Error: 0.00  
Seq primer: -40RP from Glbco  
High quality sequence stop: 273  
POLYA-No.

FEATURES  
source Location/Qualifiers

1..327  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl004-3214"  
/clone\_1lb="Gm-cl004"  
/issue\_type="root"  
/lab\_host="XL10-Gold"  
/note="Vector: pBluescript II Xr; Site\_1: EcoRI; Site\_2: XhoI; Root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. StrataGene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. StrataGene's first-strand synthesis primer was used (GAGACAGACAGACAGACTACTCTCGAG(T)-18). After second-strand synthesis, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GIBCOBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into StrataGene's pBluescript II Xr Predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies 9n-15) have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax: 520-523-7500, email: [paul.keim@nau.edu](mailto:paul.keim@nau.edu),

BASE COUNT 97 a 73 c 63 g 94 t  
 ORIGIN

Query Match 56.0%; Score 14; DB 20; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ctcgcagatgcag 25  
 |||||  
 Db 188 CTCGTCAGTCGACG 201

RESULT 23  
 C39749/c 360 bp mRNA EST 18-OCT-1999  
 LOCUS C39749 yuji kohara unpublished cDNA:Strain N2 hermaphrodite embryo  
 DEFINITION Caenorhabditis elegans CDNA clone yk217c3 5', mRNA sequence.  
 ACCESSION C39749  
 VERSION C39749.1 GI:2375986  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea  
 ; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 360)  
 Kohara, Y., Morishashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano  
 M., Miyata, A. and Nishigaki, A.  
 Expression map of the C.elegans genome  
 Unpublished (1996)  
 CONTACT: Yuji Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 81-559-81-6854  
 Fax: 81-559-81-6855  
 Email: ykohara@lab.nig.ac.jp.  
 Location/Qualifiers  
 1. 360  
 /organism="Caenorhabditis elegans"  
 /strain="N2"  
 /db\_xref="taxon:6239"  
 /clone="yk217c3"  
 /clone\_lib="Yuji Kohara unpublished cDNA:Strain N2  
 hermaphrodite embryo"  
 /sex="hermaphrodite"  
 /dev\_stage="embryo"  
 /dev\_stage="embryo"

BASE COUNT 116 a 71 c 114 g 58 t 1 others  
 ORIGIN

Query Match 56.0%; Score 14; DB 156; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cattcgatcctcg 15  
 |||||  
 Db 105 CATTGATCCTCG 92

RESULT 24  
 AM699577/c 368 bp mRNA EST 18-APR-2000  
 LOCUS AM699577  
 DEFINITION g008t06.y1 Moss EST library PPN Physcomitrella patens CDNA clone  
 PPN\_SOURCE\_ID:PPN081112 5', mRNA sequence.  
 ACCESSION AM699577  
 VERSION AM699577.1 GI:7583670  
 KEYWORDS EST.  
 SOURCE Physcomitrella patens.  
 ORGANISM Physcomitrella patens.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
 Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.

## REFERENCE

1 (bases 1 to 368)  
 Quatrano, R., Bashardes, S., Cove, D., Cumming, A., Knight, C., Clifton  
 S., Marra, M., Hillier, L., Pape, D., Martin, J., Mylie, T., Underwood  
 R., Theising, B., Allen, M., Bowers, Y., Peterson, B., Swaller, T.,  
 Steptoe, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McEam, R.,  
 Waters, S., R. and Wilson, R.  
 Leeds/Mash U Moss EST Project  
 Unpublished (1999)  
 CONTACT: Ralph Quatrano  
 Leeds/Mash U Moss EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

Libraries were constructed by Dr. Stavros Bashardes as part of the  
 Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and  
 Washington Univ. in St. Louis (USA) DNA sequencing by: Washington  
 University Genome Sequencing Center for information on obtaining a  
 clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)  
 Seq primer: -40RP from gibco.

## FEATURES

1. 368  
 Location/Qualifiers  
 /organism="Physcomitrella patens"  
 /db\_xref="taxon:3218"  
 /clone="PPN\_SOURCE\_ID:PPN081112"  
 /clone\_lib="Moss EST library PPN"  
 /lssue\_type="protonemata: 7 day old tissue auxin treated"  
 /lab\_host="DH10B"  
 /note="Vector: pluescript SK-; Site:1: EcoRI; Site:2:  
 XhoI; Construction of the CDNA library was carried out  
 using Stratagene's 'UniZAP - CDNA synthesis kit'. CDNA was  
 constructed using an oligo dT primer/linker that contains  
 a XhoI site within it. Following ds cDNA synthesis,  
 EcoRI adapters were ligated to the blunt ends and sample  
 was digested with XhoI. The result is cDNA with an EcoRI  
 sticky end on one side and a XhoI sticky end on the other.  
 This cDNA was ligated directionally in UniZAP arms. The  
 vector is designed containing the pluescript sequence as  
 well as lambda DNA and cDNA is cloned within this  
 pluescript sequence. The vector was then packaged using  
 Gold gigaPackaging extracts. Library was grown in XLBlue  
 MRF cells and amplified. The library was excised by mass  
 excision using Stratagene's 'Mass excision kit' that uses  
 exsist as a helper phage that releases the pluescript  
 sequence and circularises it as single stranded plasmids  
 that are then packaged (by helper phage) and secreted out  
 of the host cell as phagemids. SOLR cells were transformed  
 with phagemids and the library was plated out on LB-amp  
 plates to select for transformants. Approximately 1,000  
 ,000 colonies were grown and recovered. The double  
 stranded plasmid library was recovered by using Qiagen  
 Midi prep kit. 2 micro grams of each library were used to  
 transform DH10B cells by electroporation."

BASE COUNT 81 a 65 c 98 g 124 t  
 ORIGIN

Query Match 56.0%; Score 14; DB 119; Length 368;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gatccctgcagct 21  
 |||||  
 Db 105 GATCCCTGCAGCT 92

RESULT 25  
 AL370675/c 378 bp mRNA EST 03-AUG-2000  
 LOCUS AL370675/c  
 DEFINITION M8A39C04H1 MBA Medicago truncatula CDNA clone MBA39C04 T7, mRNA  
 sequence.  
 ACCESSION AL370675

VERSION AL370675.1 GI:9670428  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.  
 REFERENCE 1 (bases 1 to 378)  
 AUTHORS Journel,E.P., Crespeau,H., van-Tuinen,D., Guzy,J., Jallion,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., GlaninaZZi-Pearson,V. and Gamas,P.  
 TITLE Medicago truncatula ESTs from nitrogen-starved roots  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Genoscope  
 Genoscope  
 Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Contact : Pascal Gamas and Etienne-Pascal Journel, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : M-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).  
 Location/Qualifiers  
 1..378  
 /organism="Medicago truncatula"  
 /cultivar="Jemalong"  
 /db\_xref="taxon:3880"  
 /clone="MBA39C04"  
 /clone\_1ib="MBA"  
 /tissue\_type="root tips"  
 /dev\_stage="harvested after 3 days of N-starvation"  
 /note="Vector: pBluescript pSK; Site\_1: EcoRI; Site\_2: XhoI; Plants were grown in an aeroponic chamber for 14 days on nitrogen-rich medium followed by 3 days on N-free medium. RNA was extracted from root tips (1-3 cm). cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zapR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using EXASit helper phage and propagated in SOUR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."  
 BASE COUNT 132 a 60 c 75 g 111 t  
 ORIGIN  
 Query Match 56.0%; Score 14; DB 105; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 atccctgcaggtg 22  
 |||||  
 Db 41 ATCCTCGTCAGGTG 28  
 RESULT 26  
 FR0035439 381 bp DNA GSS 22-OCT-1999  
 LOCUS Fugu rubripes GSS sequence, clone 018f14ad12, genomic survey  
 DEFINITION sequence.  
 ACCESSION AL122961  
 VERSION 1  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Takifugu rubripes.  
 ORGANISM Takifugu rubripes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.  
 REFERENCE 1 (bases 1 to 381)  
 AUTHORS Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K., Umranta,Y., Williams,G. and Brenner,S.

TITLE Direct Submission  
 JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB, UK Email:  
 COMMENT bihelpe@mp.mrc.ac.uk  
 Vector: pBluescript II KS  
 V-type: phagemid  
 PRIMER: KS  
 DESCR: One pass dye-terminator sequencing of cosmid cloned genomic sequence.  
 Location/Qualifiers  
 1..381  
 /organism="Takifugu rubripes"  
 /db\_xref="taxon:31033"  
 /clone\_1ib="cosmid 018f14"  
 /clone="018f14ad12"  
 BASE COUNT 81 a 83 c 87 g 127 t 3 others  
 ORIGIN  
 Query Match 56.0%; Score 14; DB 222; Length 381;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 tcgacatcgatcag 19  
 |||||  
 Db 148 TCGATCGTCGTCAG 161  
 RESULT 27  
 BE021456 420 bp mRNA EST 21-NOV-2000  
 LOCUS sm49c06.y1.Gm-c1028 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1028-7235 5' similar to TR:022555 022555 O-METHYLRANSFERASE. 12  
 DEFINITION 1 TR:022308 ;, mRNA sequence.  
 ACCESSION BE021456  
 VERSION BE021456  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.  
 REFERENCE 1 (bases 1 to 420)  
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Corvelli,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Rilter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
 TITLE Public Soybean EST Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Shoemaker R/Public soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
 Insert length: 1255 Std Error: 0.00  
 High quality sequence stop: 286.  
 Location/Qualifiers  
 1..420  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1028-7235"  
 /clone\_1ib="Gm-c1028"  
 /tissue\_type="roots of 'superpod' plants"



# FEATURES

Location/Qualifiers  
1. 429

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Plate-326 Col=24 Row=J"  
/sex="male"  
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 120 a 84 c 71 g 153 t 1 others  
ORIGIN

Query Match 56.0%; Score 14; DB 225; Length 429;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 cctcgtcagtgca 24  
Db 338 CCTCGTCAGTGCA 325

RESULT 30  
LOCUS AI272538 430 bp mRNA EST 18-NOV-1998  
DEFINITION uk05h06.y1 Schiller mouse MAC13 Mus musculus cDNA clone  
IMAGE:195855 5' similar to gb:S37431 LAMININ RECEPTOR (HUMAN);  
gb:J02870 Mouse laminin receptor mRNA, complete cds (MOUSE);, mRNA  
sequence.

ACCESSION AI272538  
VERSION AI272538.1 GI:3894806  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 430)  
AUTHORS Maria M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,  
Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,  
Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,  
Theising B., Wylie T., Lennon G., Soares B., Wilson R., and  
Waterston R.

TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Maria M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:984895

Trace considered overall poor quality  
Seq primer: Primer name ambiguous  
High quality sequence stop: 1.

## FEATURES

source

1. 430  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="Schiller mouse MAC13"  
/tissue\_type="colon cancer"  
/cell\_line="colon cancer cell line MAC13"  
/lab\_host="SOLR"

/note="Vector: pBluescript SK- (Stratagene); Site\_1: EcoRI  
; Site\_2: XhoI; Double-stranded cDNA was prepared from  
cell line MAC13 using primer  
5'-GAGAGAGAGAGAGAGAGAACTGCTGAGT(18)-3'. An EcoRI  
adaptor was used on the 5' end of the cDNA as follows:  
5'-ATTGCGCAGCAG-3'. The library was size-selected and  
went through one round of amplification. Average insert

size is 1.7 kb, with a range from 0.4-12 kb. This library  
was constructed by Dr. Martin Schiller (Johns Hopkins  
University).

BASE COUNT 122 a 116 c 91 g 101 t  
ORIGIN

Query Match 56.0%; Score 14; DB 18; Length 430;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gattcctcagagt 21  
Db 80 GATCCTCGTCAGT 67

RESULT 31  
LOCUS A0129442 452 bp DNA GSS 23-SEP-1998  
DEFINITION HS-3045\_A1-A05\_MF CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=3045 Col=9 Row=A, DNA sequence.  
ACCESSION A0129442  
VERSION A0129442.1 GI:3506608  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 452)  
AUTHORS Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,  
Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,  
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and  
Hood L.

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589

CONTACT: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3045 row: A column: 9  
Class: BAC ends  
High quality sequence stop: 452.

Location/Qualifiers  
1. 452

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Plate=3045 Col=9 Row=A"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 119 a 99 c 81 g 153 t  
ORIGIN

Query Match 56.0%; Score 14; DB 224; Length 452;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 attcgtcctcgt 16  
Db 398 ATTGCGTCCTCGT 385

RESULT 32  
LOCUS AK017623 454 bp mRNA HTC 08-FEB-2001  
DEFINITION Mus musculus 8 days embryo cDNA, RIKEN full-length enriched

library, clone:5730437C11, full insert sequence.  
 AK017623  
 VERSION AK017623.1 GI:12856959  
 KEYWORDS CAP trapper.  
 SOURCE Mus musculus (strain:C57BL/6J) 8 days embryo cDNA to mRNA, clone\_11b:RIKEN full-length enriched mouse cDNA library clone:5730437C11.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (sites)  
 AUTHORS Carninci,P. and Hayashizaki,Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Methods Enzymol. 303, 19-44 (1999)  
 REFERENCE 2 (sites)  
 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 REFERENCE 3 (sites)  
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuina,T., Tashiro,H., Itoh,M., Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Mitsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913  
 REFERENCE 4 (sites)  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 MEDLINE 5 (bases 1 to 454)  
 REFERENCE 6 (bases 1 to 454)  
 AUTHORS Aachidi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Harada,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Konda,M., Koya,S., Kirihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-JUL-2000) Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, gfl:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)  
 COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to R0 = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGACGATTAAATTAATATCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pluscript KS(+) after bulk excision from Lambda F10 I. Cloning sites, 5' end: SalI; 3'

end: BamHI, Host: DH10B.  
 FEATURES  
 source Location/Qualifiers  
 1..454  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /db\_xref="MGD:MGI:1902961"  
 /db\_xref="MGD:MGI:1917808"  
 /clone\_11b="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="8 days embryo"  
 BASE COUNT 127 a 98 c 131 g 98 t  
 ORIGIN  
 Query Match 56.0%; Score 14; DB 192; Length 454;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 12 ctcgtcaggtcagc 25  
 |||||||||  
 Db 24 CTCGTCAGGTGCGAC 11  
 RESULT 33  
 A0815087  
 LOCUS  
 DEFINITION HS\_5261\_B1\_C08\_S66 RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=837 Col=15 Row=F, DNA sequence.  
 ACCESSION A0815087  
 VERSION A0815087.1 GI:5776065  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 473)  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
 Plate: 837 row: F column: 15  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 473.  
 FEATURES  
 source Location/Qualifiers  
 1..473  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_11b="RPCI-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"



BASE COUNT 132 a 99 c 124 g 116 t 2 others  
 ORIGIN

Query Match 56.0%; Score 14; DB 234; Length 473;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 cctgcagcagtgca 24  
 |||  
 Db 367 CCTGCAGCAGTGCA 380

RESULT 34  
 A0066422 486 bp mRNA EST 27-NOV-2000  
 LOCUS A0066422 Cryptomeria japonica inner bark Cryptomeria japonica cDNA  
 DEFINITION clone CC0116, mRNA sequence.  
 ACCESSION A0066422.1 GI:11370771  
 VERSION EST.  
 KEYWORDS Japanese cedar.  
 SOURCE Cryptomeria japonica  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Cupressaceae;  
 Cryptomeria.  
 REFERENCE 1 (bases 1 to 486)  
 AUTHORS Ujino, T.  
 TITLE Analysis of expressed sequence tags derived from inner bark of  
 Cryptomeria japonica  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Tokuko Ujino  
 Bio-resources Technology Division  
 Forestry and Forest Products Research Institute  
 Matsunosato 1, Kukizaki, Ibaragi 305-8687, Japan  
 Tel: +81-298-73-3211  
 Fax: +81-298-73-3795  
 Email: udino@ffpri.affrc.go.jp  
 Ujino, T. Single strand conformation polymorphism(SSCP) analysis of  
 STS markers based on partial sequences of cDNA clones in  
 Cryptomeria japonica  
 Ujino, T., Yoshimura, K., Tsumura, Y., Iwata, H., Yoshimaru, H. and  
 Nagasaka, K. Direct Submission.  
 FEATURES  
 source  
 1. 486  
 /organism="Cryptomeria japonica"  
 /db\_xref="taxon:3369"  
 /clone="CC0116"  
 /clone\_lib="Cryptomeria japonica inner bark"  
 /tissue\_type="inner bark"  
 BASE COUNT 126 a 108 c 126 g 123 t 3 others  
 ORIGIN

Query Match 56.0%; Score 14; DB 107; Length 486;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ttcgacccctgca 18  
 |||  
 Db 27 TTGCATCTCTGCTCA 40

RESULT 35  
 A1507864 494 bp mRNA EST 18-APR-2000  
 LOCUS A1507864  
 DEFINITION sa8b04.y1 Gm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 Gm-cl004-6344 5' similar to TR:022308 O22308 7-O-METHYLTRANSFERASE.  
 [3] TR:022309 TR:022555 ;, mRNA sequence.  
 ACCESSION A1507864.1 GI:4405715  
 VERSION EST.  
 KEYWORDS soybean.  
 SOURCE

ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.  
 REFERENCE 1 (bases 1 to 494)  
 AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelidng, J., Coryell, V., Khanna  
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
 Wylie, T., Underwood, K., Steptoe, M., Rheising, B., Allen, M., Bowers  
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk  
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann  
 R., Waterston, R. and Wilson, R.  
 TITLE Public Soybean EST Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 Insert length: 1307 Std Error: 0.00  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 393  
 POLYA=NO.  
 FEATURES  
 source  
 1. 494  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl004-6344"  
 /clone\_lib="Gm-cl004"  
 /tissue\_type="root"  
 /lab\_host="XL10-Gold"  
 /note="Vector: pBluescript II Xr; Site: 1: EcoRI; Site: 2:  
 XhoI; Root cDNA. The mRNA was isolated from entire roots  
 of 8 day old 'Williams' seedlings which were propagated on  
 paper towels with distilled water. StrataGene's cDNA  
 Synthesis Kit (catalog #200401) was used to synthesize the  
 cDNA. First-strand synthesis was performed with 5-methyl  
 dCTP, hence the ligated cDNA is hemimethylated.  
 StrataGene's first-strand synthesis primer was used  
 (GACAGCAGACGACGACACTGCTCTGAG(T)-18). After  
 second-strand synthesis, the cDNA ends were 'polished'  
 with clone pfu DNA polymerase, ligated to EcoRI adapters,  
 and phosphorylated. The XhoI site within the first-strand  
 synthesis primer was restricted by digestion with XhoI;  
 all XhoI sites in the cDNA would be protected by their  
 hemimethylated status. The cDNA constructs were  
 size-fractionated with a 500bp cutoff, using GibcoBRL Life  
 Technologies' cDNA size fractionation column. The column  
 eluent was then ligated into StrataGene's pBluescript II  
 XR Predigested vector (pBluescript II SK(+)) that had been  
 digested with EcoRI and XhoI, and phosphorylated). Both  
 the white and blue colonies appear to contain recombinant  
 plasmids with cDNA inserts. Blue colonies (n=15) have been  
 sequenced, and possess putative cDNA inserts. This library  
 was constructed by Dr. Paul Keim & Virginia H. Coryell,  
 Department of Biology, Box6640, Northern Arizona  
 University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.  
 Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:  
 520-523-7500, email: paul.keim@nau.edu,  
 virginia.coryell@nau.edu"

BASE COUNT 144 a 103 c 101 g 145 t 1 others  
 ORIGIN

Query Match 56.0%; Score 14; DB 21; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 ctgcgtcaggtgcag 25  
 |||||  
 Db 211 CTCGTCAGGTGAG 224

RESULT 36  
 LOCUS BF007983 518 bp mRNA EST 06-OCT-2000  
 DEFINITION 1619492 Amblyomma americanum adult Lambda Zap Express Amblyomma  
 americanum cDNA, mRNA sequence.  
 ACCESSION BF007983  
 VERSION BF007983.1 GI:10708258  
 KEYWORDS EST.  
 SOURCE Amblyomma americanum.  
 ORGANISM Amblyomma americanum.  
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
 Parasitiformes; Ixodida; Ixodidae; Amblyomma.  
 1 (bases 1 to 518)  
 Hill, C.A. and Gutierrez, J.A.  
 Analysis of the expressed genome of the lone star tick, Amblyomma  
 americanum (Acari: Ixodidae) using an expressed sequence tag  
 approach  
 Microb. Comp. Genomics 5 (2), 89-101 (2000) In press  
 Contact: Hill CA  
 Animal Science Discovery Research  
 Elianco Animal Health, A Division of Eli Lilly and Company  
 PO Box 708, 2001 West Main St., Greenfield, IN 46140, USA  
 Tel: 317 277 0826  
 Fax: 317 277 4522  
 Email: HILL.CATHERINE.A@LILLY.COM.

FEATURES  
 source  
 1. 518  
 location/Qualifiers  
 /organism="Amblyomma americanum"  
 /db\_xref="taxon:6943"  
 /clone\_lib="Amblyomma americanum adult Lambda Zap Express"  
 /sex="Male, Female"  
 /dev\_stage="Adult"  
 /note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 69 a 164 c 148 g 134 t 3 others

ORIGIN

Query Match 56.0%; Score 14; DB 143; Length 518;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 ctgcgtcaggtgcag 25  
 |||||  
 Db 27 CTCGTCAGGTGAG 40

RESULT 37  
 LOCUS AL379335 521 bp mRNA EST 03-AUG-2000  
 DEFINITION MCB44F11R1 MCBR Medicago truncatula cDNA clone MCB44F11 T7, mRNA  
 sequence.  
 ACCESSION AL379335  
 VERSION AL379335.1 GI:9679087  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.  
 1 (bases 1 to 521)  
 Journet, E.P., Crespeau, H., van Tuinen, D., Guzy, J., Jaillon, O.,  
 Niebel, A., Carreau, V., Chataigner, O., Kahn, D., Glanina, Z., Pearson,  
 V., and Gamas, P.  
 Medicago truncatula ESTs from Sinorhizobium meliloti-induced root  
 nodules  
 unpublished (2000)  
 JOURNAL Contact: Genoscope

Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr Web: www.genoscope.cns.fr  
 Contact: Pascal Gamas and Etienne-Pascal Journet, Laboratoire de  
 Biologie Moleculaire des Relations Plantes-Microorganismes,  
 CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email:  
 Mt-est@toulouse.inra.fr Website:  
 http://sequence.toulouse.inra.fr/Mtruncatula.html).

FEATURES  
 source  
 location/Qualifiers  
 1. 521  
 /organism="Medicago truncatula"  
 /cultivar="Jemalong"  
 /db\_xref="taxon:3880"  
 /clone\_lib="MtrB44F11"  
 /clone\_lib="MtrB"  
 /tissue\_type="symbiotic root nodules"  
 /dev\_stage="harvested 4 days post inoculation with  
 Sinorhizobium meliloti"  
 /note="Vector: pRiuescript pSK; Site\_1: EcoRI; Site\_2:  
 XhoI; Plants were grown in an aeroponic chamber on  
 nitrogen-rich medium for 21 days. Three days before  
 inoculation with Sinorhizobium meliloti, the medium was  
 replaced by N-free medium. Root nodules (+ short adjacent  
 root segments) were harvested 4 days post inoculation.  
 cDNA was prepared from polyA+ enriched RNA. The cDNA was  
 directionally ligated into Uni-zap XR vector from  
 Strategene and packaged using Gigapack Gold packaging  
 extracts. Plasmids containing cDNA inserts were  
 mass-excised from phage stocks using Exsist helper phage  
 and propagated in SOLR cells. Clone ordering and  
 sequencing was performed by the Centre National de  
 Sequencage (Genoscope, Evry, France)."

BASE COUNT 168 a 92 c 104 g 157 t

ORIGIN

Query Match 56.0%; Score 14; DB 105; Length 521;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atccctcaggtcag 22  
 |||||  
 Db 186 ATCTCTGTCAGGTG 173

RESULT 38  
 LOCUS BF634261 524 bp mRNA EST 19-DEC-2000  
 DEFINITION NF084G10DPI1F1083 Drought Medicago truncatula cDNA clone NF084G10DP  
 5', mRNA sequence.  
 ACCESSION BF634261  
 VERSION BF634261  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.  
 1 (bases 1 to 524)  
 Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,  
 Flores, H.R., Imman, J.T., Weller, J.W. and May, G.D.  
 Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
 Medicago truncatula drought library  
 unpublished (2000)  
 Contact: May GD  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7391  
 Fax: 580 221 7380  
 Email: gdmay@noble.org  
 Insert length: 524 Std Error: 0.00  
 Plate: 084 row: G column: 10

FEATURES  
source

Seq primer: TCACACAGGAACAGCTATCAGC.

Location/Qualifiers

1. 524

/organism="Medicago truncatula"

/db\_xref="taxon:3880"

/clone="MF084G10DP"

/clone\_lib="Drought"

/tissue\_type="Plantlets"

/dev\_stage="Pooled timepoints"

/note="Vector: Lambda Zap; Contains a mixture of entire plantlets harvested in a series of days-post-watering timepoints."

timepoints."

BASE COUNT 133 a 126 c 102 g 162 t 1 others

ORIGIN

Query Match 56.0%; Score 14; DB 151; Length 524;

Best Local Similarity 100.0%; Pred. No. 1,1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 atccctcagtg 22

Db 524 ATCCCTCAGCTG 511

RESULT 39

CNS04RAN/C

LOCUS

DEFINITION

003119 of library H from Tetradon nigroviridis, genomic survey

sequence.

AL303656.1

VERSION

KEYWORDS

SOURCE

ORGANISM

Accession

AL303656.1

GI:8187784

GSS; genome survey sequence.

Tetradon nigroviridis.

Tetradon nigroviridis.

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorphae; Acanthopterygii; Percormorpha; Tetraodontiformes;

Tetraodontidae; Tetraodon.

1 (bases 1 to 532)

Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,

Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and

Weissenbach,J.

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetradon nigroviridis

Unpublished

2 (bases 1 to 532)

Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,

Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.

Human gene number estimate provided by genome wide analysis using

Tetradon nigroviridis DNA sequence

Unpublished

3 (bases 1 to 532)

Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,

Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.

Genoscope.

Direct Submission

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetradon nigroviridis

genome. For more information, please take a look at

http://www.genoscope.cns.fr/Tetradon.

Location/Qualifiers

1. 532

/organism="Tetradon nigroviridis"

/db\_xref="taxon:99883"

/clone="003119"

/clone\_lib="H"

/note="Genoscope sequence ID : COBH003AE10XD1-end : T7"

BASE COUNT 140 a 108 c 138 g 144 t 2 others

ORIGIN

Query Match 56.0%; Score 14; DB 221; Length 532;

Best Local Similarity 100.0%; Pred. No. 1,1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 ttccatccctgc 17

Db 506 TTTCGATCCGCTGC 493

RESULT 40

AL381376

LOCUS

DEFINITION

03-AUG-2000

AL381376/c

sequence.

Accession

AL381376

KEYWORDS

SOURCE

ORGANISM

Accession

AL381376

GI:9681127

EST.

barrel medic.

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

1 (bases 1 to 548)

Journel,E.P., Crespeau,H., Van-Puinen,D., Gouzy,J., Jallion,O.,

Niebel,A., Carreau,V., Chataigner,O., Kahn,D., Gianinazzi-Pearson

,V. and Gamas,P.

Medicago truncatula ESTs from endomycorrhizal roots

Unpublished (2000)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Contact : Pascal Gamas and Etienne-Pascal Journel, Laboratoire de

Biologie Moleculaire des Relations Plantes-Microorganismes,

CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :

mt-est@toulouse.inra.fr Website :

http://sequence.toulouse.inra.fr/Mtruncatula.html).

Location/Qualifiers

1. 548

/organism="Medicago truncatula"

/cultivar="Jemalong"

/db\_xref="taxon:3880"

/clone="MCBC019F0"

/clone\_lib="MCBC"

/tissue\_type="arbuscular mycorrhiza"

/dev\_stage="harvested 3 weeks post inoculation with Glomus

intraradices"

/note="Vector: pBluescript PSK; Site\_1: EcoRI; Site\_2:

XhoI; M. truncatula sterilised seeds were germinated for

72h at 25 C, before transplanting into a 1/3 Eposides soil

; 2/3 calcined terragreen mix in the presence of onion

root fragments colonized by the arbuscular mycorrhizal

fungus Glomus intraradices (Schwenk & Smith, isolate LPB8

). The plants were watered every day and twice a week with

a modified nutrient Long Ashton solution without phosphate

but with a high level of nitrate. After 3 weeks RNA was

extracted from whole root systems. cDNA was prepared from

polyA+ enriched RNA. The cDNA was directionally ligated

into Uni-zap XR vector from Stratagene and packaged using

GigaPack Gold packaging extracts. Plasmids containing cDNA

inserts were mass-excised from phage stocks using Exsist

helper phage and propagated in SOUR cells. Clone ordering

and sequencing was performed by the Centre National de

Sequencage (Genoscope, Evry, France). Note : EST may be of

fungal origin.

BASE COUNT 168 a 101 c 112 g 167 t

ORIGIN

Query Match 56.0%; Score 14; DB 105; Length 548;

Best Local Similarity 100.0%; Pred. No. 1,1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atccctgctcagtg 22  
 |||  
 Db 179 ATCTGCTGCTGAGTG 166

RESULT 41  
 A0528128/c  
 LOCUS A0528128/c  
 DEFINITION RPT1-11-372K15.TV RPT1-11 Homo sapiens genomic clone RPT1-11-372K15  
 / DNA sequence.  
 ACCESSION A0528128  
 VERSION A0528128.1 GI:4840282  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 559)  
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.

TITLE Use of BAC End Sequences from Library RPT1-11 for Sequence-Ready  
 Map Building  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbe@tigr.org  
 Clones are derived from the human BAC library RPT1-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@edlong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
 Research Genet cs ([inforesgen.com](http://inforesgen.com)). BAC end search page:  
[http://www.tigr.org/tldb/humgen/Bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/humgen/Bac_end_search/bac_end_search.html).  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
 source  
 location/Qualifiers  
 1..559  
 /organism="Homo sapiens"  
 /db\_xref="CDB:7642718"  
 /db\_xref="taxon:9606"  
 /clone="RPT1-11-372K15"  
 /clone\_lib="RPT1-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: PBAC3.6; Site:1: ECORI; Site-2: ECORI;  
 RPT11 Human Male BAC Library"  
 BASE COUNT 115 a 150 c 151 g 141 t 2 others  
 ORIGIN

Query Match 56.0%; Score 14; DB 229; Length 559;  
 Best Local Similarity 100.0%; Pred. No. 1.le+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 cctcgtcagtgca 24  
 |||  
 Db 374 CCTGCTGCTGCTGCA 361

RESULT 42  
 BG080788/c  
 LOCUS BG080788/c  
 DEFINITION H3056F11-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone  
 H3056F11 5', mRNA sequence.  
 ACCESSION BG080788  
 VERSION BG080788.1 GI:12563279  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 586)  
 AUTHORS Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka,T.S., Carter,M.G. and Ko,M.S.H.  
 TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set  
 JOURNAL Unpublished (2001)  
 COMMENT Other-ESTs: H3056F11-3  
 Contact: George J. Kargul  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6620, USA  
 Email: cdna@igsun.grc.nia.nih.gov  
 This clone set has been freely distributed to the community. Please  
 visit <http://igsun.grc.nia.nih.gov/cDNA/15k.html> for details.  
 Plate: H3056 row: F column: 11  
 Seq primer: -21M13 Reverse  
 High quality sequence stop: 586  
 POLYA-No.

FEATURES  
 source  
 location/Qualifiers  
 1..586  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="H3056F11"  
 /clone\_lib="NIA Mouse 15K cDNA Clone Set"  
 /sex="Clones arrayed from a variety of cDNA libraries"  
 /dev\_stage="Clones arrayed from a variety of cDNA libraries"  
 /lab\_host="DH10B"  
 /note="Vector: pSPOR1; Site:1: SalI; Site-2: NotI. This  
 clone is among a rearranged set of 15,247 clones from 11  
 embryo cDNA libraries (including preimplantation stage  
 embryos from unfertilized egg to blastocyst, embryonic  
 part of E7.5 embryos, extraembryonic part of E7.5 embryos  
 , and E12.5 female mesonephros/gonad) and one newborn  
 ovary cDNA library. Average insert size 1.5 kb. All  
 source libraries are cloned unidirectionally with Oligo(dT  
 )-Not primers. References include: (1) Genome-wide  
 expression profiling of mid-gestation placenta and embryo  
 using a 15,000 mouse developmental cDNA microarray, 2000,  
 Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)  
 Large-scale cDNA analysis reveals phased gene expression  
 patterns during preimplantation mouse development, 2000,  
 Development, 127: 1737-1749; (3) Genome-wide mapping of  
 unselected transcripts from extraembryonic tissue of  
 7.5-day mouse embryos reveals enrichment in the t-complex  
 and under-representation on the X chromosome, 1998, Hum  
 Mol Genet 7: 1967-1978."

BASE COUNT 189 a 120 c 121 g 156 t  
 ORIGIN

Query Match 56.0%; Score 14; DB 173; Length 586;  
 Best Local Similarity 100.0%; Pred. No. 1.le+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gatcctcagtg 21  
 |||  
 Db 547 GATCTCTGCTGAGT 534

RESULT 43  
 A0655910/c  
 LOCUS A0655910/c  
 DEFINITION Sheared DNA-817.TF Sheared DNA Trypanosoma brucei genomic clone  
 A0655910  
 ACCESSION A0655910  
 VERSION A0655910.1 GI:5163678  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.

```

REFERENCE      1 (bases 1 to 610)
AUTHORS        El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
                Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
                Fraser,C. and Adams,M.
TITLE          Determination of clone end sequences from Trypanosoma brucei GUTat
                10.1 sheared DNA library
JOURNAL        Unpublished (1999)
COMMENT        Other_GSS: Sheared DNA-817.JR
                Contact: Najib M. El-Sayed
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: neisayed@tigr.org
                Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
                DNA library constructed at TIGR. Clones will be available for
                distribution through ATCC. Sheared DNA end sequences search page:
                http://www.tigr.org/tldb/mbd/tbdb/.
                Seq primer: M13-Forward
                Class: Shotgun.

FEATURES
source         Location/Qualifiers
                1..610
                /organism="Trypanosoma brucei"
                /strain="TREU927/4 GUTat 10.1"
                /db_xref="taxon:5691"
                /clone="Sheared DNA-817"
                /clone_1lb="Sheared DNA"
                /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                Institute for Genomic Research (TIGR), Rockville, MD.
                Genomic DNA isolated from a cloned population of
                Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                sheared to give a tight size distribution (approx 2 kb).
                The v + i method used for the library construction is
                described in detail in Smith, H.O. and Venter, J.C.
                (Making small insert libraries for whole genome shotgun
                sequencing projects. In Genome Sequencing: A Practical
                Approach, eds. M. Vaudin and B. Barrell, Oxford University
                Press, 1999)."

BASE COUNT     145 a      125 c      194 g      146 t

ORIGIN
Query Match    56.0%; Score 14; DB 231; Length 610;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY             9 atctctgcacgtg 22
                |||||||
Db             433 ATCTCTGTCACGTCG 420

RESULT 44
AM953979      642 bp      mRNA      EST      01-JUN-2000
LOCUS         EST365944 MAGE resequences, MAGC Homo sapiens cDNA, mRNA sequence.
DEFINITION    AM953979
ACCESSION     AM953979
VERSION       AM953979.1 GI:8143557
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 642)
Hege,P., Qi,R., Abernathy,K., Dharap,S., Gaspari,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source         Location/Qualifiers
                1..642
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_1lb="MAGE resequences, MAGC"
                /note="Vector: pBluescriptSKm"

BASE COUNT     166 a      151 c      174 g      151 t

ORIGIN
Query Match    56.0%; Score 14; DB 122; Length 642;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY             9 atctctgcacgtg 22
                |||||||
Db             354 ATCTCTGTCACGTCG 341

RESULT 45
BG067650      644 bp      mRNA      EST      26-JAN-2001
LOCUS         H3056F11-3 NIA Mouse 15K cDNA clone set Mus musculus cDNA clone
DEFINITION    H3056F11 3', mRNA sequence.
ACCESSION     BG067650
VERSION       BG067650.1 GI:12550219
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 644)
Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
,T.S., Carter,M.G. and Ko,M.S.H.
Verification and initial annotation of NIA mouse 15K cDNA clone set
Unpublished (2001)
Other_ESTs: H3056F11-5
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.igr.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://19sun.igr.nia.nih.gov/cDNA/15k.html for details.
Plate: H3056 row: F column: 11
Seq primer: -21M13 Forward
High quality sequence stop: 644
POLY(A)-yes.

FEATURES
source         Location/Qualifiers
                1..644
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="H3056F11"
                /clone_1lb="NIA Mouse 15K cDNA Clone Set"
                /sex="Clones arrayed from a variety of cDNA libraries"
                /dev_stage="Clones arrayed from a variety of cDNA
                libraries"
                /lab_host="DH10B"
                /note="Vector: pSPORI1; Site_1: SalI; Site_2: NotI; This
                clone is among a rearranged set of 15,247 clones from 11
                embryo cDNA libraries (including preimplantation stage
                embryos from unfertilized egg to blastocyst, embryonic
                part of E7.5 embryos, extraembryonic part of E7.5 embryos
                and E12.5 female mesonephros/gonad) and one newborn
                ovary cDNA library. Average insert size 1.5 kb. All
                source libraries are cloned unidirectionally with Oligo(dT
                )-NotI primers. References include: (1) Genome-wide

```

expression profiling of mid-gestation placenta and embryo  
using a 15,000 mouse developmental cDNA microarray, 2000,  
Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)  
large-scale cDNA analysis reveals phased gene expression  
patterns during preimplantation mouse development, 2000,  
Development, 127: 1737-1749; (3) Genome-wide mapping of  
unselected transcripts from extraembryonic tissue of  
7.5-day mouse embryos reveals enrichment in the t-complex  
and under-representation on the X chromosome, 1998, Hum  
Mol Genet 7: 1967-1978."

BASE COUNT 239 a 96 c 104 g 205 t  
ORIGIN

Query Match 56.08; Score 14; DB 173; Length 644;  
Best Local Similarity 100.08; Pred. No. 1.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gatcctcgtcaggt 21  
|||||  
Db 581 GATCCTCGTCAGGT 594

Search completed: October 9, 2001, 15:15:37  
Job time: 13659 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:47:46 ; Search time 1670.83 Seconds  
(without alignments)  
231.438 Million cell updates/sec

Title: US-09-396-196f-9

Sequence: 1 cgatccctcgtcagtgacagtcagc 25

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 205018

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_om:\*  
9: gb\_om:\*  
10: gb\_pat1:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_inv3:\*  
33: em\_htg\_inv4:\*  
34: em\_htg\_rod:\*  
35: em\_hum1:\*  
36: em\_hum2:\*  
37: em\_hum3:\*  
38: em\_hum4:\*  
39: em\_hum5:\*  
40: em\_hum6:\*  
41: em\_hum7:\*  
42: em\_in:\*  
43: em\_om:\*  
44: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_rod:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_rol:\*  
95: gb\_rol2:\*  
96: gb\_rol3:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID      | Description        |
|------------|-------|-------------|--------|------------|--------------------|
| 1          | 25    | 100.0       | 1041   | 9 AR029499 | AR029499 Sequence  |
| 2          | 25    | 100.0       | 1041   | 9 AR034916 | AR034916 Sequence  |
| 3          | 25    | 100.0       | 1084   | 9 A11530   | A11530 B10B gene o |
| 4          | 25    | 100.0       | 1121   | 10 E00893  | E00893 Genomic DNA |
| 5          | 25    | 100.0       | 5793   | 2 ECOBIO   | J04423 E.coli 7,8- |
| 6          | 25    | 100.0       | 5872   | 9 A38246   | A38246 Sequence 1  |
| 7          | 25    | 100.0       | 5872   | 9 A38251   | A38251 Sequence 6  |
| 8          | 25    | 100.0       | 5872   | 9 A93674   | A93674 Sequence 1  |

```

9      25 100 0 5872 9 A93679
10     25 100 0 5872 9 ARI01809
11     25 100 0 5872 9 ARI01810
12     25 100 0 11022 1 AE000180
13     25 100 0 13501 1 AE005258
14     25 100 0 297816 2 AE002553
15     18 72 0 7215 3 SMABIO
16     17 68 0 7318 94 AB026497
17     17 68 0 69900 66 AC021767
18     17 68 0 151340 74 AC069132
19     17 68 0 155204 73 AC068100
20     17 68 0 168360 71 AC041048
21     17 68 0 176379 72 AC060833
22     17 68 0 200729 67 AC022781
23     16 64 0 10536 1 AE004050
24     16 64 0 71230 85 AC004508
25     16 64 0 110000 84 HSMX1_4
26     16 64 0 155628 71 AC027810
27     16 64 0 183700 86 AC005412
28     16 64 0 213732 1 AE001862
29     16 64 0 313064 93 HSMX1B
30     16 64 0 340000 92 HS21C085
31     15 60 0 128 2 ECOBIOB
32     15 60 0 636 88 AF072252
33     15 60 0 728 15 TC057884
34     15 60 0 1552 8 XLHNP4MR
35     15 60 0 1727 85 AB056427
36     15 60 0 1801 14 RICCTYALD
37     15 60 0 2081 94 HAMHP
38     15 60 0 2931 8 DRAJ5029
39     15 60 0 3829 93 HSMETHYL7
40     15 60 0 4674 88 AF034373
41     15 60 0 4886 85 AB037816
42     15 60 0 5556 2 AF250776
43     15 60 0 8326 8 AF237379
44     15 12542 1 AE004708
45     15 60 0 20780 63 AC014847

```

## ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
    source
        Location/Qualifiers
            1..1041
                /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 cgatctcgtcaggtgcaggtcagc 25
    |||||||
Db 105 CGATCCTCGTCAGTGACAGTCAGC 129

```

```

AR034916
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
    source
        Location/Qualifiers
            1..1041
                /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgatctcgtcaggtgcaggtcagc 25
    |||||||

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Patent: GB 2216530-A 16 11-OCT-1989;
JOURNAL Location/Qualifiers
FEATURES
    source
        Location/Qualifiers
            1..1084
                /organism="Escherichia coli"
                /db_xref="taxon:562"
            24..1064
                /gene="B10B"
            24..1064
                /gene="B10B"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA00965.1"
            /db_xref="GI:490219"
            /db_xref="SWISS-PROT:P12996"
            /translation="MAHRRRWITLSQYTELEFKRPLDLLEAQQVHRQHPDPOVOYST
            LLSIKGACPEDCKTPQSSRIKTYGLEARLEVEEVLBSARKAKAAGSTRCKGAW
            KNPHERDMPLEQWVGVRKAMGLEACMTIGTISESOARLANAGLDYVNNHNDTSPEF
            YGNIITTRVYQERLDYLEKVRDAGIVCGSGIVGIGETVKDRAGLILQLANIPPTPES
            VPINMLVKVGTPLADNDVDVADFRTIYAVARIIMPSTSYVRLSAGREQOMOTQAMC
            FMAGANSIFYGCKILTPNPBDEKDLQLEFRKILGLNQQRAVALAGNEQQRLQALMT
            PTDDEVYNNAAAL"
            /protein_id="P12996"
            /db_xref="SWISS-PROT:P12996"
            /translation="MAHRRRWITLSQYTELEFKRPLDLLEAQQVHRQHPDPOVOYST
            LLSIKGACPEDCKTPQSSRIKTYGLEARLEVEEVLBSARKAKAAGSTRCKGAW
            KNPHERDMPLEQWVGVRKAMGLEACMTIGTISESOARLANAGLDYVNNHNDTSPEF
            YGNIITTRVYQERLDYLEKVRDAGIVCGSGIVGIGETVKDRAGLILQLANIPPTPES
            VPINMLVKVGTPLADNDVDVADFRTIYAVARIIMPSTSYVRLSAGREQOMOTQAMC
            FMAGANSIFYGCKILTPNPBDEKDLQLEFRKILGLNQQRAVALAGNEQQRLQALMT
            PTDDEVYNNAAAL"
BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgatctcgtcaggtgcaggtcagc 25
    |||||||

```





```

/codon_start=1
/transl_table=11
/protein_id="AA23517.1"
/db_xref="GI:145427"
/translation="MATVKNKOATAAAGRAAHYEOHADLOROSADALLAMPORRYT
HVIDAGGPGMRSRRHRRERAAQYATLDSLPMIVOROKDAADHYLAGDIESPLATA
TFDLANSNLAVOMCGNSTRLRELRYVRPKGVAFITLVQSLPERHQAOWADVERP
HANRFLPPDEIEQSLNGVHYQHIIQPIITLWFDDALSAMSLKIGATHLHEGRDRL
TRSLQRLQALAMPQOOGRYPLTYHFLGVYARE"
4938..5597
/gene="bioc"
/gene="bioc"
4938..5597
/gene="bioc"
/codon_start=1
/transl_table=11
/product="delhiolotin synthetase"
/protein_id="AA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVNTDTEVEKTYASCALLQAAKACRTIRAGYRPVASGS
EKPEEKLMSDALAORNSLQIDTAVNPYFAEPTSPHLSAQGRPIESLMSAG
LRLEQOQADWVLVEGAGWFTPLSDTFADWVTOEQPLVLVGVKLGICINHAMLTA
QVIOHAGLTLAGVANDVTPGKRHAHYMTTLTMRIPRCWERSPGLOKIQKMQPES
T"
BASE COUNT      1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.00066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      cgatccctgcagtgtagtcagc 25
Db      2116  CGATCCCTGCTGAGTGAGTGCAGC 2140
|||||
RESULT      6
LOCUS      A38246      5872 bp      DNA
DEFINITION      Sequence 1 from Patent WO9408023.
ACCESSION      A38246
VERSION      A38246.1      GI:2294844
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch, O., Brass, U., Fuhrmann, M. and Shaw, N.
TITLES      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL      Patent: WO 9408023-A 1 14-APR-1994;
COMMENT
LONZA AG (CH)
other publication PL 308301 950724
other publication AU 2145400 940414
other publication CA 2145400 940426
other publication HU 4820293 940426
other publication HU 71781 960228
other publication SK 42095 951108
other publication CZ 9500809 950913
other publication FI 951547 950331
other publication JP 8501694T 960227.
FEATURES
source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PHO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
/evidence=experimental
45..49
/standard_name="PROMOTER PTAC"
/evidence=experimental

RBS
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/number=1
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="MAHRPRWLTLSQVETLEPEKPLDLLFPAQOYHROHPPOVYST
LLSTKTCGPEPDCYKQSSRYKTGLEABERLMEYOEYLSARRAKAGSTRFCGAAM
KNPHRDMPTLEQWVOGVKAMGLEACWTLGLTSSQAKRLANGLDYNNILPTSPES
YGNITTKTYOERLDTLEKVDAGIKVCSGIVGLGTVDKAGLLIQLANLPTPEPS
VPINMLVKQCTPLADNDVDADPFIITAVARIIMPTSYVRLSAGREOMNEOTOAMC
PMAGANSIFYCKLITTPNPEDEKDLQLFRLGLINPOQTAVLAGDNEQOQLRQALMT
PDTEYVNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMETOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVKNKOATAAAGRAAHYEOHADLOROSADALLAMPORRYT
HVIDAGGPGMRSRRHRRERAAQYATLDSLPMIVOROKDAADHYLAGDIESPLATA
TFDLANSNLAVOMCGNSTRLRELRYVRPKGVAFITLVQSLPERHQAOWADVERP
HANRFLPPDEIEQSLNGVHYQHIIQPIITLWFDDALSAMSLKIGATHLHEGRDRL
TRSLQRLQALAMPQOOGRYPLTYHFLGVYARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOMATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDLAPDORHIMHPYTSMTSPILPVYVSAEGCELLISGR
LYNGSSMAAIIHGVNHPOLNAAMKSDIDAMSHVWGTTTAPALBELCKVLAMPPOP
LELVFLADSGSAVAVAMKALQYWOAKSEAKQRLTFRNGYHGDFFGMSVCDPDS
MISLWKGYLEPENLEFAPQSRMDGEMDERDNGFARLMAAHHEIAVAITIEPIVQAGS
GMRMYHPMLTKRIRKICDREGIILIADEIATGFGRTGKLFACENHAIEDILCKAL
TGCTWTLISATLITREVAETISNGEAGCFPMHGTFPMGNPLACAAANASIALILSGDMOQ
QVADIEVOLREOLAPARDAEMVADVILGALIGVETTHPVYNAALOKPFVEGQWIRP
FGKLIIMPYIILPQOLQRLTAANVRAVQDETFPCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

```
/evidence-experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRDGLKPHRHVFNKGMDGNISPLIAMDVPAGTK
SEVVTGYPDPAFGSGMHWYVNLPAETRYLPQFSGSLVAMPDGYLQPTDQKIG
YDGAAPRGERTHRTFTYVHALDIERIDVDEGASGAMGFNVHIFSLASATITAFS"
5583..5603
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match
Best local similarity 100.0%; Score 25; DB 9; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 cgatcctgcaggtgcagtcagc 25
        |||||||
Db      221 CGATCCTGCTCAGGTGCAGTCAGC 245

RESULT 7
LOCUS      A38251      5872 bp      DNA      PAT      05-MAR-1997
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION  A38251
VERSION     A38251.1  GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL    Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     other publication PL 308301 950724
            other publication CA 2145400 940414
            other publication AU 4820293 940426
            other publication HU 71781 960228
            other publication SK 42095 951108
            other publication CZ 9500809 950913
            other publication FI 951547 950331
            other publication JP 8501694T 960227.
            Location/Qualifiers
                1..5872
                /organism="Escherichia coli"
                /strain="DSM498"
                /db_xref="taxon:562"
                /clone="PB030A15-9"
                1141..1156
                /standard_name="BIOF RBS"
                1154..2308
                /gene="BIOF"
                1154..2308
                /gene="BIOF"
                /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                /EC_number="2.3.1.47"
                /codon_start=1
                /transl_table=11
                /number=2
                /evidence-experimental
                /product="KAPA SYNTHASE"
                /protein_id="CAA02329.1"
                /db_xref="GI:2294850"
                /translation="MSWDERINAAALDARAADALRRRYVAGACGMLVADDRQYLNF
                SSNDYGLSHHPQIIIRAWOGAEQFGIGSGSCHVSGVVAHQALEELAEWIGYSRA
                LEFISGFANQAVITAAAKEDRIADRLSHASLSLEASISPSQLRFAHNDVTHLAR
                LLASPCGQOMVTEGVFSMDGSAPFLAEIQVYTOHNGMLVMDAHRGTGIVGEQGR
```

```
SCWLOKVPPELLVTFGKGVSAAVLCSSTVADYLLQFAHLLYTSMPQAQAL
RASLAVIRSEGDARREKIALITRRAGVODLPFLADSCASIOPLIYVGDNSRALQL
AEKLRQGGCWVTAIRPPYPAGTARLRLLTAHMQIDRLLEVIHNG"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence-experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFVTGDTFEVGTVASCALLQNAKAGYRTAGYKPVASGS
EKTPGELRNSDALAIORNSSLQDYAVTNVYFAEPTSPHIIISAGRPRESLVMASG
LRALFQADAVLVHAGAGGFTPLSDPTTFADWYTOROLPVILVGVKIGCINRAMLTA
QVTOHAGLTLGWAYANDVTPPGKRHAETMTTLRMLPAPLGEIPLAENPERNAATGK
YINLAFVDASTLFTSRL"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match
Best local similarity 100.0%; Score 25; DB 9; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 cgatcctgcaggtgcagtcagc 25
        |||||||
Db      221 CGATCCTGCTCAGGTGCAGTCAGC 245

RESULT 8
LOCUS      A93674      5872 bp      DNA      PAT      22-JAN-2000
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION  A93674
VERSION     A93674.1  GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O. and Brass, J.
TITLE      Biotechnological method of producing biotin
JOURNAL    Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
COMMENT     Location/Qualifiers
                1..5872
                /organism="Escherichia coli"
                /strain="DSM498"
                /db_xref="taxon:562"
                /clone="PB030A-15/9"
                1..96
                /function="PROMOTER PTAC"
                /evidence-experimental
                23..28
                /standard_name="PROMOTER PTAC"
                45..49
                /evidence-experimental
                105..119
                /standard_name="BIOB RBS NO.9"
                /evidence-experimental
                117..1157
                /gene="BIOB"
                117..1157
                /gene="BIOB"
                /codon_start=1
```



BASE COUNT 1318 a 1552 c 1695 g 1307 t  
 YINAFVDASTLGFSTRU"  
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 0.00065;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cgaatcctcgtcaggtcaggtcagc 25  
 |||||||  
 Db 221 CGATCCTCGTCAGGTGCGAGGTGACG 245

RESULT 10  
 ARI01809 ARI01809 5872 bp DNA PAT 14-FEB-2001  
 LOCUS Sequence 1 from patent US 6083712.  
 ACCESSION ARI01809  
 VERSION ARI01809.1 GI:12812607  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)  
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.  
 TITLE Biotechnological method of producing biotin  
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;  
 FEATURES Location/Qualifiers  
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t  
 ORIGIN /organism="unknown"

Query Match 100.0%; Score 25; DB 9; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 0.00065;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cgaatcctcgtcaggtcaggtcagc 25  
 |||||||  
 Db 221 CGATCCTCGTCAGGTGCGAGGTGACG 245

RESULT 11  
 ARI01810 ARI01810 5872 bp DNA PAT 14-FEB-2001  
 LOCUS Sequence 6 from patent US 6083712.  
 ACCESSION ARI01810  
 VERSION ARI01810.1 GI:12812608  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)  
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.  
 TITLE Biotechnological method of producing biotin  
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;  
 FEATURES Location/Qualifiers  
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t  
 ORIGIN /organism="unknown"

Query Match 100.0%; Score 25; DB 9; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 0.00065;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cgaatcctcgtcaggtcaggtcagc 25  
 |||||||  
 Db 221 CGATCCTCGTCAGGTGCGAGGTGACG 245

RESULT 12  
 AE000180 AE000180 11022 bp DNA BCT 01-DEC-2000  
 LOCUS Escherichia coli K12 MG1655 section 70 of 400 of the complete  
 DEFINITION genome.

ACCESSION AE000180 000096  
 VERSION AE000180.1 GI:1786988  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE 1 (bases 1 to 11022)  
 AUTHORS Escherichia coli K12.  
 Escherichia coli K12  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.

1 (bases 1 to 11022)  
 Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,  
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,  
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,  
 Mau, B. and Shao, Y.  
 The complete genome sequence of Escherichia coli K-12  
 Science 277 (5331), 1453-1474 (1997)  
 97426617  
 9278503

2 (bases 1 to 11022)  
 Blattner, F.R.  
 Direct Submission  
 Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,  
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:  
 608-263-7459

3 (bases 1 to 11022)  
 Blattner, F.R.  
 Direct Submission  
 Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,  
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:  
 608-263-7459

4 (bases 1 to 11022)  
 Plunkett, G. III.  
 Direct Submission  
 Submitted (13-OCT-1998) Laboratory of Genetics, University of  
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 This sequence was determined by the E. coli Genome Project at the  
 University of Wisconsin-Madison (Friedrick R. Blattner, director).  
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome  
 Project and NCHGP). The entire sequence was independently  
 determined from E. coli K12 strain MG1655. Predicted open reading  
 frames were determined using Genemark software, kindly supplied by  
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,  
 30332 [e-mail: markborov@gatech.edu]. Open reading frames that  
 have been correlated with genetic loci are being annotated with CG  
 Site Nos., unique ID nos. for the genes in the E. coli Genetic  
 Stock Center (CGSC) database at Yale University, kindly supplied by  
 Mary Berlyn. A public version of the database is accessible  
 (http://cgsc.biology.yale.edu). Annotation of the genome is an  
 ongoing task whose goal is to make the genome sequence more useful  
 by correlating it with other data. Comments to the authors are  
 appreciated. Updated information will be available at the E. coli  
 Genome Project's World Wide Web site  
 (http://www.genetics.wisc.edu). \*\*\* The E. coli K12 sequence and  
 its annotations are periodically updated; this is version M54. No  
 sequence changes. Annotation updates: updated gene identifications  
 and products; all new functional assignments courtesy of Monica  
 Riley; added promoters, protein binding sites, and repeated  
 sequences described in reference 1. The unique numeric identifiers  
 beginning with a lowercase 'b' assigned to each gene (protein- or  
 RNA-encoding) are now designated as gene synonyms instead of  
 labels. This should allow them to be searched for in Entrez as gene  
 names.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

FEATURES  
 source

1..11022  
 /organism="Escherichia coli K12"  
 /strain="K12"  
 /sub\_strain="MG1655"

repeat\_region  
 <1..130  
 /db\_xref="taxon:83333"  
 /note="REP (repetitive extragenic palindromic) element; contains 4 REP sequences"

gene  
 complement(147..1430)  
 /gene="ybhc"  
 /note="b0772"  
 complement(147..1430)  
 /gene="ybhc"  
 /function="orf; Not classified"  
 /note="f427; 98 pct identical to fragment YBHC\_ECOLI SW:P46130 (300 aa) but contains 127 additional C-terminal residues"

CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative pectinesterase"  
 /protein\_id="AAC73859.1"  
 /db\_xref="GI:1786989"  
 /translation="MNTFVSRLALAFGLTLAGSSPPDQPSDQRTAGTSSRL  
 LSAKQNPFAOHVFASTLPGAAANPSPPTLPAODPVPYGPACTGCTHTT1QAAD  
 AALIKRTNRQYIAVMPGEYQGVYVPAAGITLGTGKPLDVKIGLSDGMSPA  
 DKRDVAPKRGIMPGKPAWYMDSCSKRSDSLGVLCSAVFMSQNGLOLQNT  
 LGSDVAGNHPAAVALRTDQVOINNVNIGRONTFPTVNSGVONRLTNRPRLVT  
 NSYIEGVADIVSGCAVNDTEFRVNSRTQOEAVFAPATLSNITYGFLAVNSRD  
 AFDGVAGLRSIDVADANTGOVYIRDSAINEGENTAKPWADAVISNRPAGVTGSVD  
 DNDEIQRNLNDTVNRKMEYNRSGSKVVAEAKK"  
 complement(1507..1535)  
 /note="factor Sigma70; predicted +1 start at 806574"  
 complement(1582..2058)  
 /gene="ybhc"  
 /note="b0773"  
 complement(1582..2058)  
 /function="orf; Unknown"  
 /note="f158; 99 pct identical to YBHC\_ECOLI SW: P12994"  
 /transl\_table=11  
 /codon\_start=1  
 /product="orf, hypothetical protein"  
 /protein\_id="AAC73860.1"  
 /db\_xref="GI:1786990"  
 /translation="MKLISNDLRDGKRLPHRHVFNMGYGDGNIISPHLAMDVPAGTK  
 SEVVTYDPAIPGSGMHVNVNLPADTRVLPQGGSLVAMPDGLDTRDFGTG  
 YDGAAPKGETHRYIPTVALDIERIDVDEGSAWVGFVHSHIASASITAMS"  
 complement(2108..2124)  
 /note="central position to predicted promoter:85.5"  
 /bound\_moiety="Rhas predicted site"  
 complement(2117..3406)  
 /gene="bioA"  
 /note="b0774"  
 complement(2117..3406)  
 /gene="bioA"  
 /EC\_number="2.6.1.62"  
 /function="enzyme; Biosynthesis of cofactors, carriers: Biotin"  
 /note="f429; 100 pct identical to BIOA\_ECOLI SW: P12995"  
 /codon\_start=1  
 /transl\_table=11  
 /product="7,8-diaminopelargonic acid synthetase"  
 /protein\_id="AAC73861.1"  
 /db\_xref="GI:1786991"  
 /translation="MTTDDLADPDRHIWHPYNSMTSPLEVPVVSAGCECLISDGR  
 LVDGMSWMAALHGYNHDPOLNAAMKSOIDAMSHVMGGITTHAPATELCRLVAMPPOP  
 LDCVPLDSSGSAVEAMKALQYAMKAEARORFLFRNGYHGTGEGMSVCDPDS  
 GMRVHGYLPENLPAPOSMDGEMDERDVGAFRLMAAHRHEIAAYITIPYOGAL  
 GSRVHPEWLRKIRKIDREGILLADELAGFGTGLFCEHAEIADIPDICGKAL  
 NGTMTLSATITTTREVAETISNGEAGCTRHGTEFGNPLACAAASIALIESGWOQ  
 QYADIEVOLREQLAPARDAENYADVRLGALGVETTHPVNMAALQKFFVBOGWIRP  
 FEKLIYLPMPYIILPQOLQRLTAAVNRAVODETFPCQ"  
 complement(2193..2221)  
 /gene="bioA"  
 /note="factor Sigma70; predicted +1 start at 807260"  
 3411..3450  
 /note="central position to bioB promoter: -20"

promoter  
 protein\_bind

protein\_bind  
 3411..3450  
 /bound\_moiety="BioB documented site"  
 /note="central position to predicted promoter: -20; genetic evidence for the site"

protein\_bind  
 /bound\_moiety="BioB predicted site"  
 complement(3411..3450)  
 /note="central position to predicted promoter:50"  
 /bound\_moiety="BioB predicted site"  
 complement(3411..3450)  
 /note="central position to bioA promoter:50"  
 /bound\_moiety="BioB documented site"  
 3413..3441  
 /note="factor Sigma70; promoter bioB; documented +1 at 808525"  
 complement(3447..3473)  
 /note="factor Sigma70; promoter bioA; documented +1 at 808515"

promoter  
 3493..4533  
 /gene="bioB"  
 /note="b0775"  
 3493..4533  
 /gene="bioB"  
 /EC\_number="2.8.1.-"  
 /function="enzyme; Biosynthesis of cofactors, carriers: Biotin"  
 /note="o346; 99 pct identical to BIOB\_ECOLI SW: P12996"  
 /codon\_start=1  
 /transl\_table=11  
 /product="biotin synthesis, sulfur insertion?"  
 /protein\_id="AAC73862.1"  
 /db\_xref="GI:1786992"  
 /translation="MAHRPWTLSQVTELEFKPRLDLLPEAQVHRQHPDQVYST  
 LSTKTVGACPEDCYKCPQSSRYKTGLEERLMEYOYLESAKKAAGSTRFCGAW  
 KNPERMDPYLEOVQGVKAMGLACMTLGLTSSQARLANAGLDYNNILDPSPF  
 YGNITTRTYOERIDTLEKVDAGIKVSGSIGVIGETVDRAGLLLOLNPSPS  
 VPINMLVVKGTPLANDNDVDAFPFIRIYAVRIMPTSVRLSAGSEONNEQOAMC  
 PMAGNSITFVGCKLITTPNPEBDLQFLKGLINPQOTAVIAGDNEQORLEQALMT  
 PPTDEYVNAAL"  
 4530..5684  
 /gene="bioB"  
 /note="b0776"  
 4530..5684  
 /gene="bioB"  
 /EC\_number="2.3.1.47"  
 /function="enzyme; Biosynthesis of cofactors, carriers: Biotin"  
 /note="o384; 100 pct identical to BIOF\_ECOLI SW: P12998"  
 /codon\_start=1  
 /transl\_table=11  
 /product="8-amino-7-oxononanoate synthase"  
 /protein\_id="AAC73863.1"  
 /db\_xref="GI:1786993"  
 /translation="MSWQEKINALDARRAADLRRRYVPAOGAGRWLVADROYLNF  
 SSNDYGLSHHPQILIRAMQGAQBOFGISGSGSHVSGVSYVHOLEBEELAEWYGSRA  
 LLFTSGPAANDAVTAAMAKEDRIARARLSHAILSAKSLSPGLRFRFANDVTYHAR  
 LLAISPCGQGVAVTEGVSMDCGSAPIALTOQVTOOHNGLMDDAHGRTVTEQGRG  
 SCHLQKPKELVYTFKRGVSGAAVLCSTVADYILQFARHLITSTMPRAQOAL"

gene  
 4530..5684  
 /gene="bioB"  
 /note="b0776"  
 4530..5684  
 /gene="bioB"

CDS  
 4530..5684  
 /gene="bioB"

Query Match 100.0%; Score 25; DB 1; Length 11022;  
 Best local Similarity 100.0%; Pred. No. 0.00059;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctgtaagtgccagcagc 25  
 |||||||  
 Db 3597 CGATCTCTGTCAGGTGCGAGTCAGC 3621

RESULT 13  
 AEO05258 13501 bp DNA BCT 21-MAR-2001  
 LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82  
 DEFINITION of 155.  
 ACCESSION AEO05258 AEO05174

| VERSION  | KEYWORDS   |
|--|--|
| AE005258.1   | GI:12513751  |
| SOURCE   | ORGANISM   |
| Escherichia coli O157:H7 EDL933.   | Escherichia coli O157:H7 EDL933  |
| Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.  |  |
| REFERENCE  | AUTHORS  |
| 1 (bases 1 to 13501)   | Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoustis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R. |
| Nature 409 (6819), 529-533 (2001)  |  |
| TITLE  | JOURNAL  |
| Genome sequence of enterohaemorrhagic Escherichia coli O157:H7   |  |
| MEDLINE  | PUBMED   |
| 21074935   | 11206551   |
| REFERENCE  | AUTHORS  |
| 2 (bases 1 to 13501)   | Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoustis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R. |
| Direct Submission  |  |
| Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  |  |
| FEATURES   | SOURCE   |
| Location/Qualifiers  | 1..13501   |
| /organism="Escherichia coli O157:H7 EDL933"  | /strain="EDL933"   |
| /serotype="O157:H7"  | /db_xref="taxon:155864"  |
| /note="enterohemorrhagic"  | <1..7576   |
| /note="O-island #36; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933K; includes one copy of the 13 bp direct repeat that flanks the prophage"                   | 66..665  |
| /gene="lomk"   | /note="lomk"   |
| /note="lomk"   | 66..665  |
| /gene="lomk"   | /function="putative membrane; Other or unknown (Phage or Prophage Related)"  |
| /note="Residues 1 to 199 of 199 are 72.86 pct identical to gi17532789 gb IAAF63231.1 AF151091_2 (AF151091) Lom [prophage P-E1bA]"  | /codon_start=1   |
| /transl_table=1  | /product="putative outer membrane protein of prophage CP-933K"   |
| /protein_id="AAG55137.1"   | /db_xref="GI:12513752"   |
| /translation="MRKVCALILSAATCIAGVGPANASEHOSLTASGYLHASTDVG SDDLNGINVKRYREFTDTGLITFSFVANADBDKTKYSPDRHEPYRRNMEYSVAG PSVAVNENFSAIAMAGYASRVSTFSGDYFVTNKKRTHDVLTGSDARKSNTSLAM GAGVQFPNTESSAADVAVAYEXSGSDWRTDGFIVGYSK" | 719..2041  |
| /gene="20982"  | 719..2041  |
| /gene="20982"  | /function="putative structure; Structural component (Phage or Prophage Related)"   |
| /note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from Genpept 118 : gi14584361 gb IAAD25464.1 AF125520.59 (AF125520) putative tail fiber protein [Bacteriophage 933W]"  | /codon_start=1   |
| /transl_table=1  | /product="putative tail component of prophage CP-933K"   |

gene  
CDS  
/protein\_id="AAG55138.1"  
/db\_xref="GI:12513753"  
/translation="MGMAAVOISGVLRKDGAKPIQNCITQLKARNSTYVVNVTVASE  
NNDGRCRSMVEYEGOYVLTLLVDEGPPPSHAGTISYEDNSOGTINDLTGMTADBAR  
PALRFRDQVVEEARHAEAKKRAKALETISANAGISAKRAEAAADAPASDAEE  
SARQAAEENSAASAKSEESSSSASAEQAQRASSETLSQATIDKRTASAGNARBA  
TSTKRAEENSSASQASABOSQIAEDAVNNITPVVAPPKKEPPAPQPKDKDKE  
RDTDPACATGACRGCGDTGAGPGGPGKGDGRGRTGTGLTGAAGQPGKGTGAAGA  
GGQGGKGGGAAGPVGATPGGPKGDPGGEFQIIFRLGPMRIETISYCMFPTGALIL  
TGLTFLDKDATQVGMQHLQVRFSGDGPMDYVKGLDEVGSDTGRTGE"  
2004..2312  
/gene="20984"  
2004..2312  
/gene="20984"  
/function="orf; Other or unknown (Phase or Prophage  
Related)"  
/note="Residues 1 to 102 of 102 are 98.03 pct identical to  
residues 28 to 129 of 129 from GenPept 118 :  
g114583437|gb|AA025465.1|AF125520\_60 (AF125520)  
hypothetical protein [Bacteriophage 933K]"  
/codon\_start=1  
/transl\_table=11  
/product="unknown protein encoded by prophage CP-933K"  
/protein\_id="AAG55139.1"  
/db\_xref="GI:12513754"  
/translation="MKAAVIOALELNDNNMLRLKMSGLCGCKHDCNGSITAAQLR  
LGAADILEDENGIILPEODRVTQVTLIDADKKQIQCVVRPIQILRADGTENIGAKM  
"  
2489..3469  
/gene="20985"  
2489..3469  
/gene="20985"  
/function="orf; Other or unknown (Phase or Prophage  
Related)"  
/note="Residues 13 to 326 of 326 are 55.73 pct identical  
to residues 18 to 331 of 336 from GenPept 118 :  
g116960367|gb|AA033527.1|AF170176) hypothetical protein  
predicted by Glimmer [Salmonella typhimurium LT2]"  
/codon\_start=1  
/transl\_table=11  
/product="unknown protein encoded by prophage CP-933K"  
/protein\_id="AAG55140.1"  
/db\_xref="GI:12513755"  
/translation="MLSPRTPTFNHSNVIYQSSPCQTSFPAKEVELAKYIDKPTLIL  
QMEPRPEPKYKDEVDYNTQOHPLQNDVNTAAATIESDRMIGLFFVGDGVSNOQAF  
KLERDENNVITREDVDPSMTDKRLSDYVHDIICBQRLPTEDKRDYELLNLKEEL  
ETSRKADSSLMITAKRNNHAWDFPRNLALKEAGEITRCYTNKNGISTFSGGCTYLIL  
MDMILVIGKGTQIYADGISMHVDRNDSVNTENSAITVNSNRPALLBGLSPHBSKV  
AHNYVDGLCKGVKAKYFTPLHNYNHPCDIEFNHPNIIIMNTSYQTCSSW"  
3530..4522  
/gene="20986"  
3530..4522  
/gene="20986"  
/function="orf; Other or unknown (Phase or Prophage  
Related)"  
/note="No significant matches"  
/codon\_start=1  
/transl\_table=11  
/product="unknown protein encoded by prophage CP-933K"  
/protein\_id="AAG55141.1"  
/db\_xref="GI:12513756"  
/translation="MKITPSLQSNFNSAPAGSYAPLAPNRAENAVADYVLIDGKRIIP  
SADLISVESVETIRAVHDSRSLIDQHTVMDIGMTVLDALSRSQTRDAVSGTINHEIP  
VHIGCIKYRNEYELNEESSVKTDIDISTLCMELKLEYVGOGEPIPFCEAGENDNEBP  
VSEVADDTSYEMPMGOBGLIHEITIHVTSSPSDSNIMELGPIETILARRVADQV  
MSYDPEGVGAVEREHLRLRNALROAARHRENNRAPERFRTGISDYEASPPFE  
EYSVANSIGTGFIQOHDFFGLINDNDANOIDLHGACIYITFPDGVKHNQO"  
5350..6231  
/gene="20989"  
5350..6231  
/gene="20989"  
/function="orf; Other or unknown (Phase or Prophage  
Related)"

| Query Match           | Score 25   | DB 1              | Length 13501 |
|-----------------------|--|-------------------|--------------|
| Best local Similarity | 100.0%   | Prid. No. 0.00057 |              |
| Matches 25            | Conservative 0   | Mismatches 0      | Indels 0     |
| Gaps 0                |  |                   |              |
| Db 9683               | CGATCCTCGTCAGTGACGATCAGC   | 9707              |              |
| Qy 1                  | cgatcctcgatcagatgacgagtcagc  | 25                |              |
|                       |  |                   |              |
| RESULT 14             |  |                   |              |
| AP002553              | 297816 bp  | DNA               |              |
| LOCUS                 |  | BCT               | 07-MAR-2001  |
| DEFINITION            | Escherichia coli O157:H7 DNA, complete genome, section 4/20.       |                   |              |
| ACCESSION             | AP002553   | BA000007          |              |
| VERSION               | AP002553.1   | GI:13360211       |              |
| KEYWORDS              | Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952) |                   |              |
| SOURCE                | DNA.   |                   |              |
| ORGANISM              | Escherichia coli O157:H7   |                   |              |
|                       | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;   |                   |              |
| REFERENCE             | 1 (sites)  |                   |              |

[illegible]



```

gene
1245..2114
/gene="ECS0754"
CDS
1245..2114
/gene="ECS0754"
/note="similar to SUCD_ECOLI g111786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSHIDKNTKVTVCQGFSGCTHSRQATVYTKMVGYPKGC
GTHLGLPENTVREAAVATGATASVIVPAPECKSDILEADIGKILITTEGIPY
LDMLTVKVIDEAGVRMIGPCGVITPGECKIGIOPGHHPKGVIVSSGTLTE
AVKOTEDGSGCVCIGDPIGNSNIDLEMEFDPQTEAIVMIGETIGSAEBA
AAVKEHTKPVVGYIAGVTPAKGKRMHAGATITAGKGTADKFAALAGAVTVRS
LADIGELKTVLK"
2781..3686
/gene="ECS0755"
2781..3686
/gene="ECS0755"
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator)"
/escherichia coli g11417043|sp|p32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPTKTELLVPEVVAHRHESYTRAAEELATGSAVEROVSAL
EEFLRTPLFSSHKRIPLNDGKYIAGIKELNLEEDNTIMTQPTVAVELAVN
PTEFHMILPNIHEFTKLPDIIVNHSIANNGDVIREYDAVIRENFCAPMAVEY
LPEERLTPVCGSLIAMSODKISVAELLTEPLHOSRTITGMEEMFALSGVSPILVN
NGPRDLISMLIAVRSNLGVALLPRAFIQHDLSGDVITCDVPIRTGNFMIMTQCE
EKSDSPHLQREEMLAKSVPOEM"
complement(3720..4322)
/gene="ECS0756"
complement(3720..4322)
/gene="ECS0756"
/note="probable cob(1)alamin adenosyltransferase, similar
to cob(1)alamin adenosyltransferases (corticoid
adenosyltransferases) e.g. [Escherichia coli]
g1115148|sp|p13040|BRUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(1)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTERRHVOROOKLKEQVDTVAALIEKKGILVFTGNGK
KSTAFGTVTRAVGCKYGVACQYIKGQMDNGEYVNLQPLGVEHFMIGTFWENOR
QADIDAKVWSESKRMILADKRYDLVYDELTYMLAYHLDVEVIASLQNRPAQSV
IYTGRCCHSQILKMDVSEIRPVKHAFFNGIQADPGIDW"
complement(4332..5984)
/gene="ECS0757"
complement(4332..5984)
/gene="ECS0757"
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarate) - Escherichia coli
g1120598|sp|p00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

```

```

/translation="MSKPIFMQEPFLONKDGTEYTLISDQHTITVELDGEYKIAPL
ALTLLARQAFYENSPFLRSAAHIQVASTIINDQASNNKYVALQILRNAAEVSAGVLP
NCDTGTATVASKGOQIWTGNDABALSKGISTFOENNLRFSONAPLDWTEYVNO
TNDLPADIDISAAVAGDEYHFLCVNKGGSNKAALQETKSLQPEKTLAFLTEKSL
GTACCPHYIAFVYGGISADQTLKIAKLASTRYYNLPTSGNEQQAQFADLEKVL
EASOQFGIAGQFVGGYFAHDIRYIRLPRRGGSCPIAMALSCSADNRKAIKNGIWL
EKIEHNPQGYIPASLRREHNAOHVOLDLRPLRDMODLARLPVGT RVSLSGIVAVR
DIHAKITKARLDGSGEPMEYIKHIIYVYAGPACTPENNMACSGISLPTGGRMGYIDTF
QANGSLVMLSKGRNSOQVTDACHKKGGRNLSIGAAALLQAEVYKSLRCLEPELG
MEAVMMEEVENLPAFLVDDKGNPFSPQEQHRCASCPAGH"
complement(6092..7372)
/gene="ECS0758"
complement(6092..7372)
/gene="ECS0758"
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g1121467|sp|p24943|GLMT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLFTMLALVGLMIGVYVNNATSPETAKLYAQEISFT
TIFRLKMTIAPLVYSTIYVIGAKMDKAKGIRSKLFLFCASLSLISLGLTV
NFMFGGTFNFVAHGAFTTVVAAPPTLVKPIFSAFPPSYVDMANHELIOVRSI
FLGSLTAIEKGSALVHALDSLAAHAKLUTGVMLPAFLPVAISALIAEGLAVM
VSAQITFMGEFFPTMLLMLVLLISLAIYVGPICRILTRALSPALLAFTTSSAEP
GLEKILRQFVSPIASFVLPICGSFVLSVMSGMAVCSFATVFIQACNINHLISIGEOTM
LTIIMLRSKMGVPRASVVVIAATLMQFNIPAGLILMGVDFLMDGRSAYNVASN
AMGAMVSRWEGEGHGGCGRKALKKPNESVALP"
complement(7533..7850)
/gene="ECS0759"
complement(7533..7850)
/gene="ECS0759"
complement(7533..7850)
/gene="ECS0759"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

Query Match 100.0% Score 25; DB 2; Length 297816;  
Best Local Similarity 100.0%; Pred. NO. 0.00034;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 cgatcctcgtcagtcagtcagtcagc 25
|||||
Db 92415 CGATCCTCCTCAGTCAGTCAGTCAGC 92439

```

| RESULT     | 15   |
|------------|--|
| LOCUS      | SMABIO 7215 bp DNA BCT 04-FEB-1999   |
| DEFINITION | S. marcescens biotin operon, complete sequence.  |
| ACCESSION  | D17468   |
| VERSION    | D17468.1 GI:402530   |
| KEYWORDS   | 7,8-diamino-pelargonic acid aminotransferase;<br>7-keto-8-aminopelargonic acid synthetase; DAPA aminotransferase;<br>DTP synthetase; KAPA synthetase; biotin operon; biotin synthetase;<br>dehydrobiotin synthetase. |
| SOURCE     | Serratia marcescens (strain:Str41) DNA.  |
| ORGANISM   | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;<br>Serratia.  |
| REFERENCE  | 1 (bases 1 to 7215)  |
| AUTHORS    | Sakurai, N.  |
| TITLE      | Direct Submission  |
| JOURNAL    | Submitted (25-AUG-1993) to the DDBJ/EMBL/Genbank databases. Naoiki<br>Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry;<br>2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan                   |

(E-mail: gfc0101@niftyserve.or.jp, nsakurai@ddbj.nig.ac.jp,  
Tel: 048-433-2545, Fax: 048-433-2540)  
2 (bases 1 to 7215)

REFERENCE  
AUTHORS  
Tosai, T., Imai, Y., Akatsuka, H., Kawai, E., Komatsu, S. and  
Tosai, T.

TITLE  
Complete nucleotide sequence of biotin operon of *Serratia*

JOURNAL  
COMMENT  
Unpublished (1993)  
Submitted (25-Aug-1993) to DDBJ by:  
Naoki Sakurai

Res. Lab. of Applied Biochemistry  
Tanabe Seiyaku Co., Ltd.  
2-50 Kawagishi-2-chome  
Toda, Saitama 335

Japan  
Phone: 048-433-2545  
E-mail: nsakurai@ddbj.nig.ac.jp  
Fax: 048-433-2540.

FEATURES  
source  
1. 7215  
Location/Qualifiers

```

/organism="Serratia marcescens"
/strain="Str41"
/db_xref="taxon:615"
/complement(1353..2630)
/gene="bioc"
/complement(1353..2630)
/gene="bioc"
/BC_number="2.6.1.62"
/codon_start=1
/transl_table=1
/product="7,8-diamino-2-pelargonic acid aminotransferase
(DAPA aminotransferase)"
/protein_id="BAA04284.1"
/db_xref="GI:402531"
/translation="MSVTRASDLAFPOORHIMPTYSMSRPLPCYPISASGVQLQADG
BSLVQSSSWMAIHGVNHPRLNOASOLEKMSHVMFGITHPAAISLCRLVEMTP
EALQCVFLADSSGVAEVSLLKALQYQWQGRGRRRIILTRHGYHGDFTGMSVCDP
NSMHSLYOGLYAPHLFATAPQCRFDEWEENEIAPFAALLEQHAQEVAVIIEPVQO
AGGMRHYPTLYLRVYDVVRAIKLLIADIELATGFGPRGKLFACENHOVVDILICG
KALTCGYMTLSATLTTRHYAERTISNGAGCFMHWGTPMGNPLACVADAASIALLENK
WQAOYSAITETOLKRELLPLGRCRRMPHRYVGAICGVAKREPVDAELQRGVEKGV
IIPFGKLIYMPYITIEBQLSRITAAVADAR"
2716..3756
/gene="biob"
/gene="biob"
2716..3756
/gene="biob"
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="BAA04285.1"
/db_xref="GI:402532"
/translation="MMADRIHMTVGOAQLFDKPLLELLFEAOTVHROHPDPOVYS
TLISITGACPEDECKYCPSPRYKTGLESPRIMOYEOVLSEARAKAKASPRFGMAA
WKNPHERMPYLOMVGVCAMGMECTGSLITLQTOQERLAEGLDIYNNIDTSE
FVGSITITRSVOERLIDTKVRDAGIKVSGGIVGLGETVDRAGLLQNLNPKRPE
SVPIMLYKVGKPLADNDVDPEDFTITAVARIIMPSSVVRUSAGREONETOAM
CFMAGASITFYGCKLLTPNPEDEKDLQLFRRKLGINPOOTATEHGMQOQVIAKOLL
NADTAEFYNAAP"
3756..4904
/gene="biob"
/gene="biob"
3756..4904
/gene="biob"
/codon_start=1
/transl_table=1
/product="7-keto-8-aminopelargonic acid synthetase (KAPA
synthetase)"
/protein_id="BAA04286.1"
/db_xref="GI:402533"
/translation="MSWOORIEQALERRRLNAAYRRQTEGGNGROIRLGDRLYLNF
SGNDLIGSDARVIAWMOGARYGVGGSGHVTGSAHQAOLEQALAEGLDIYNNIDTSE
LFISYAQAVALAALMOKGRILADRLSHSLILFAAGAVGRAPVPAOSIAGTARIC
LAKPCDQRLAVETGLTSMODGAPLAEHLRLTRAAGAMLVDDAHGICIGVREGGRS
CMQGVRELLVATEFGKAFVSGAAVLCDEATATYLLQFAHLLITYSTAMPDAOCALO

```

gene  
CDS

AALNIREGDDLKARLQDNIIRFRQAGAPLALTLTDSPTALQPLVGNDRALDLATR  
LREGLWVSARIPPTVPPGARRLRLITLAAHQSDIDRLLEVLNDVSG"

4888..5655  
/gene="bioc"  
4888..5655  
/gene="bioc"

/function="synthesis of pimeloyl-CoA"

/codon\_start=1

/transl\_table=1

/product="the product of bioc"

/protein\_id="BAA04287.1"

/db\_xref="GI:402534"

/translation="MTSANDVNVKQAVASAFSRAAGSYDAALQDRVGRLLMGSS  
HPGQLDAGCGGTGVFSRMMREGRKRVYALDLAPRMDVARORQAAHNYLLEQYV  
LPDAADICEFSLVWQMSDLPALAEIYRTVPGGYILFSTLAAGSLQELGAMQOY  
DGERVNAFLPLTQITACAAVREHLYTELTNTLPVPMMLMSKIGTATHLQGRE  
GLMSRGLAALQAAVRCRQGPFLSHLAVGYIRE"

5648..6331  
/gene="biob"

/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

5648..6331  
/gene="biob"

QY 8 cytcaggtcagtcagc 25

DB 2830 CGTCAGGTGAGTCAGC 2847

RESULT 16

AB026497/c

LOCUS

DEFINITION

AB026497

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Best Local Similarity 72.0%; Score 18; DB 3; Length 7215;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AB026497 7318 bp mRNA ROD 29-APR-2000  
Mus musculus mySPD mRNA for myosin containing PDZ domain, complete  
cds.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.

AB026497.1 GI:7416031  
Mus musculus (strain:C57BL/6) adult tissue\_11b:spleen cDNA to mRNA,  
clone\_11b:lambda Ziplox.



\* 788 887: gap of 100 bp  
\* 888 1651: contig of 764 bp in length  
\* 1652 1751: gap of 100 bp  
\* 1752 2546: contig of 795 bp in length  
\* 2547 2646: gap of 100 bp  
\* 2647 3438: contig of 792 bp in length  
\* 3439 3538: gap of 100 bp  
\* 3539 4327: contig of 789 bp in length  
\* 4328 4427: gap of 100 bp  
\* 4428 5234: contig of 807 bp in length  
\* 5235 5334: gap of 100 bp  
\* 5335 6129: contig of 795 bp in length  
\* 6130 6229: gap of 100 bp  
\* 6230 7040: contig of 811 bp in length  
\* 7041 7140: gap of 100 bp  
\* 7141 7931: contig of 791 bp in length  
\* 7932 8031: gap of 100 bp  
\* 8032 8810: contig of 779 bp in length  
\* 8811 8910: gap of 100 bp  
\* 8911 9696: contig of 786 bp in length  
\* 9697 9796: gap of 100 bp  
\* 9797 10596: contig of 800 bp in length  
\* 10597 10696: gap of 100 bp  
\* 10697 11493: contig of 797 bp in length  
\* 11494 11593: gap of 100 bp  
\* 11594 12398: contig of 806 bp in length  
\* 12400 12499: gap of 100 bp  
\* 12500 13296: contig of 797 bp in length  
\* 13297 13396: gap of 100 bp  
\* 13397 14152: contig of 756 bp in length  
\* 14153 14252: gap of 100 bp  
\* 14253 15034: contig of 782 bp in length  
\* 15035 15134: gap of 100 bp  
\* 15135 15922: contig of 788 bp in length  
\* 15923 16022: gap of 100 bp  
\* 16023 16795: contig of 773 bp in length  
\* 16796 16895: gap of 100 bp  
\* 16896 17698: contig of 803 bp in length  
\* 17699 17798: gap of 100 bp  
\* 17799 18568: contig of 770 bp in length  
\* 18569 18668: gap of 100 bp  
\* 18669 19457: contig of 789 bp in length  
\* 19458 19557: gap of 100 bp  
\* 19558 20342: contig of 785 bp in length  
\* 20343 20442: gap of 100 bp  
\* 20443 21239: contig of 797 bp in length  
\* 21240 21339: gap of 100 bp  
\* 21340 22122: contig of 783 bp in length  
\* 22123 22222: gap of 100 bp  
\* 22223 22995: contig of 773 bp in length  
\* 22996 23095: gap of 100 bp  
\* 23096 23836: contig of 741 bp in length  
\* 23837 23936: gap of 100 bp  
\* 23937 24720: contig of 784 bp in length  
\* 24721 24820: gap of 100 bp  
\* 24821 25590: contig of 770 bp in length  
\* 25591 25690: gap of 100 bp  
\* 25691 26480: contig of 790 bp in length  
\* 26481 26580: gap of 100 bp  
\* 26581 27368: contig of 788 bp in length  
\* 27369 27468: gap of 100 bp  
\* 27469 28258: contig of 790 bp in length  
\* 28259 28358: gap of 100 bp  
\* 28359 29134: contig of 776 bp in length  
\* 29135 29234: gap of 100 bp  
\* 29235 30016: contig of 782 bp in length  
\* 30017 30116: gap of 100 bp  
\* 30117 30897: contig of 781 bp in length  
\* 30898 30997: gap of 100 bp  
\* 30998 31770: contig of 773 bp in length  
\* 31771 31870: gap of 100 bp  
\* 31871 32662: contig of 792 bp in length  
\* 32663 32762: gap of 100 bp

\* 32763 33558: contig of 796 bp in length  
\* 33559 33658: gap of 100 bp  
\* 33659 34448: contig of 790 bp in length  
\* 34449 34548: gap of 100 bp  
\* 34549 35346: contig of 798 bp in length  
\* 35347 35446: gap of 100 bp  
\* 35447 36251: contig of 805 bp in length  
\* 36252 36351: gap of 100 bp  
\* 36352 37136: contig of 785 bp in length  
\* 37137 37236: gap of 100 bp  
\* 37237 38021: contig of 785 bp in length  
\* 38022 38121: gap of 100 bp  
\* 38122 38908: contig of 787 bp in length  
\* 38909 39008: gap of 100 bp  
\* 39009 39800: contig of 792 bp in length  
\* 39801 39900: gap of 100 bp  
\* 39901 40683: contig of 783 bp in length  
\* 40684 40783: gap of 100 bp  
\* 40784 41551: contig of 768 bp in length  
\* 41552 41651: gap of 100 bp  
\* 41652 42409: contig of 758 bp in length  
\* 42410 42509: gap of 100 bp  
\* 42510 43302: contig of 793 bp in length  
\* 43303 43402: gap of 100 bp  
\* 43403 44204: contig of 802 bp in length  
\* 44205 44304: gap of 100 bp  
\* 44305 45097: contig of 793 bp in length  
\* 45098 45197: gap of 100 bp  
\* 45198 45976: contig of 779 bp in length  
\* 45977 46076: gap of 100 bp  
\* 46077 46856: contig of 780 bp in length  
\* 46857 46956: gap of 100 bp  
\* 46957 47735: contig of 779 bp in length  
\* 47736 47835: gap of 100 bp  
\* 47836 48609: contig of 774 bp in length  
\* 48610 48709: gap of 100 bp  
\* 48710 49519: contig of 810 bp in length  
\* 49520 49619: gap of 100 bp  
\* 49620 50405: contig of 786 bp in length  
\* 50406 50505: gap of 100 bp  
\* 50506 51282: contig of 777 bp in length  
\* 51283 51382: gap of 100 bp  
\* 51383 52152: contig of 770 bp in length  
\* 52153 52252: gap of 100 bp  
\* 52253 53044: contig of 792 bp in length  
\* 53045 53144: gap of 100 bp  
\* 53145 53941: contig of 797 bp in length  
\* 53942 54041: gap of 100 bp  
\* 54042 54836: contig of 795 bp in length  
\* 54837 54936: gap of 100 bp  
\* 54937 55730: contig of 794 bp in length  
\* 55731 55830: gap of 100 bp  
\* 55831 56616: contig of 786 bp in length  
\* 56617 56716: gap of 100 bp  
\* 56717 57498: contig of 782 bp in length  
\* 57499 57598: gap of 100 bp  
\* 57599 58381: contig of 783 bp in length  
\* 58382 58481: gap of 100 bp  
\* 58482 59250: contig of 769 bp in length  
\* 59251 59350: gap of 100 bp  
\* 59351 60139: contig of 789 bp in length  
\* 60140 60239: gap of 100 bp  
\* 60240 61030: contig of 791 bp in length  
\* 61031 61130: gap of 100 bp  
\* 61131 61917: contig of 787 bp in length  
\* 61918 62017: gap of 100 bp  
\* 62018 62811: contig of 794 bp in length

Query Match 68.0%; Score 17; DB 66; Length 69900;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 cgtcaggtcaggtcag 24

DB 19825 CGTCAGCTCAGCTCAG 19809

|||||

RESULT 18

AC069132

LOCUS Homo sapiens chromosome 5 clone RP11-556L8, \*\*\* SEQUENCING IN

DEFINITION PROGRESS \*\*\* 62 unordered pieces.

ACCESSION

AC069132.2 GI:8469032

VERSION HTG: HTGS\_PHASE1.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

On Jun 12, 2000 this sequence version replaced gi:7924001.

\*\*\*\*\* Genome Center \*\*\*\*\*

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 62 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1502: contig of 1502 bp in length  
1503 1602: gap of unknown length  
1603 2741: contig of 1139 bp in length  
2742 2841: gap of unknown length  
2842 4036: contig of 1195 bp in length  
4037 4136: gap of unknown length  
4137 5620: contig of 1484 bp in length  
5621 5720: gap of unknown length  
5721 7080: contig of 1360 bp in length  
7081 7180: gap of unknown length  
7181 8839: contig of 1659 bp in length  
8840 8939: gap of unknown length  
8940 9987: contig of 1048 bp in length  
9988 10087: gap of unknown length  
10088 11156: contig of 1069 bp in length  
11157 11256: gap of unknown length  
11257 12404: contig of 1148 bp in length  
12405 12504: gap of unknown length  
12505 13861: contig of 1357 bp in length  
13862 13961: gap of unknown length  
13963 15241: contig of 1280 bp in length  
15242 15341: gap of unknown length  
15342 17009: contig of 1668 bp in length  
17010 17109: gap of unknown length  
17110 18897: contig of 1788 bp in length  
18898 18997: gap of unknown length  
18998 20389: contig of 1392 bp in length  
20390 20489: gap of unknown length  
20490 22042: contig of 1553 bp in length  
22043 22142: gap of unknown length  
22143 24057: contig of 1915 bp in length  
24058 24157: gap of unknown length  
24158 26189: contig of 2032 bp in length  
26190 26289: gap of unknown length  
26290 27959: contig of 1670 bp in length  
27960 28059: gap of unknown length

28060 30067: contig of 2008 bp in length  
30167 30168: gap of unknown length  
30168 32456: contig of 2289 bp in length  
32457 32536: gap of unknown length  
32536 34597: contig of 2041 bp in length  
34597 34598: gap of unknown length  
34598 36611: contig of 1914 bp in length  
36611 36712: gap of unknown length  
36712 38677: contig of 1966 bp in length  
38677 38778: gap of unknown length  
38778 40974: contig of 2197 bp in length  
40974 41074: gap of unknown length  
41074 42666: contig of 1592 bp in length  
42666 42667: gap of unknown length  
42667 44292: contig of 1526 bp in length  
44292 44392: gap of unknown length  
44392 46032: contig of 1640 bp in length  
46032 46132: gap of unknown length  
46132 48815: contig of 2683 bp in length  
48815 48915: gap of unknown length  
48915 48916: gap of unknown length  
48916 51058: contig of 2143 bp in length  
51058 51158: gap of unknown length  
51158 53094: contig of 1936 bp in length  
53094 53194: gap of unknown length  
53194 54724: contig of 1530 bp in length  
54724 54824: gap of unknown length  
54824 57019: contig of 2195 bp in length  
57019 57119: gap of unknown length  
57119 59736: contig of 2617 bp in length  
59736 59836: gap of unknown length  
59836 61510: contig of 1674 bp in length  
61510 61610: gap of unknown length  
61610 63447: contig of 1637 bp in length  
63447 63448: gap of unknown length  
63448 65262: contig of 1915 bp in length  
65262 65348: gap of unknown length  
65348 65362: gap of unknown length  
65362 67493: contig of 2131 bp in length  
67493 67593: gap of unknown length  
67593 70740: contig of 3147 bp in length  
70740 70840: gap of unknown length  
70840 72256: contig of 1416 bp in length  
72256 72356: gap of unknown length  
72356 75456: contig of 3100 bp in length  
75456 75556: gap of unknown length  
75556 77328: contig of 1772 bp in length  
77328 77428: gap of unknown length  
77428 79851: contig of 2423 bp in length  
79851 79951: gap of unknown length  
79951 82877: contig of 2926 bp in length  
82877 82977: gap of unknown length  
82977 84859: contig of 1882 bp in length  
84859 84959: gap of unknown length  
84959 87856: contig of 2897 bp in length  
87856 87956: gap of unknown length  
87956 91677: contig of 3721 bp in length  
91677 91777: gap of unknown length  
91777 94536: contig of 2759 bp in length  
94536 94637: gap of unknown length  
94637 97377: contig of 2741 bp in length  
97377 97477: gap of unknown length  
97477 100981: contig of 3504 bp in length  
100981 101081: gap of unknown length  
101081 103958: contig of 2877 bp in length  
103958 104058: gap of unknown length  
104058 106903: contig of 2845 bp in length  
106903 107003: gap of unknown length  
107003 109662: contig of 2659 bp in length  
109662 109762: gap of unknown length  
109762 113672: contig of 3910 bp in length  
113672 113772: gap of unknown length  
113772 118353: contig of 4581 bp in length  
118353 118453: gap of unknown length  
118453 121541: contig of 3088 bp in length

```

* 121542 121641: gap of unknown length
* 121642 124122: contig of 2481 bp in length
* 124123 124222: gap of unknown length
* 124223 128114: contig of 3892 bp in length
* 128115 128214: gap of unknown length
* 128215 131684: contig of 3470 bp in length
* 131685 131784: gap of unknown length
* 131785 135575: contig of 3791 bp in length
* 135576 135675: gap of unknown length
* 140587 140687: contig of 4912 bp in length
* 140688 145142: contig of 4455 bp in length
* 145143 145243: gap of unknown length
* 145243 151340: contig of 6098 bp in length.
Location/Qualifiers
1. 151340
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-556L8"
BASE COUNT 43058 a 29029 c 29780 g 43356 t 6117 others
ORIGIN

```

```

Query Match 68.0% Score 17; DB 74; Length 151340;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 cgtcaggtcaggtcag 24
|||||
Db 93019 CGTCAGGTCAGGTCAG 93035

```

```

RESULT 19
AC068100 AC068100 155204 bp DNA HTG 16-MAR-2001
DEFINITION Homo sapiens chromosome X clone RP11-812N12 map X, WORKING DRAFT
ACCESSION AC068100
VERSION AC068100.4 GI:13357469
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155204)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
JOURNAL Homo sapiens chromosome X, clone RP11-812N12
REFERENCE 2 (bases 1 to 155204)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boucknight,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choquet,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talmas,J.,
Tessier,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,R., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

```

TITLE  
JOURNAL  
Submitted (28-Apr-2000) Whitehead Institute/MIT Center for Genome

## COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 16, 2001 this sequence version replaced gi:12583851.  
All repeats were identified using RepeatMasker.  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information -----  
Center project name: L7131  
Center clone name: 812\_N12

----- Summary Statistics -----  
Sequencing vector: M13; M77815; 6% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 146541 bases at least Q40  
Consensus quality: 150817 bases at least Q30  
Consensus quality: 152368 bases at least Q20  
Insert size: 165000; agarose-fp

Quality coverage: 4.5 in Q20 bases; agarose-fp  
Quality coverage: 4.8 in Q20 ba.

NOTE: This is a 'working draft' sequence. It currently  
consists of 21 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

```

1 518: contig of 518 bp in length
* 519 618: gap of 100 bp
* 619 1733: contig of 1115 bp in length
* 1734 1833: gap of 100 bp
* 1834 2717: contig of 884 bp in length
* 2718 2817: gap of 100 bp
* 2818 5343: contig of 2526 bp in length
* 5344 5443: gap of 100 bp
* 5444 8556: contig of 3113 bp in length
* 8557 8656: gap of 100 bp
* 8657 12409: contig of 3753 bp in length
* 12410 12509: gap of 100 bp
* 12510 16818: contig of 4309 bp in length
* 16819 16918: gap of 100 bp
* 16919 18984: contig of 2066 bp in length
* 18985 19084: gap of 100 bp
* 19085 41463: contig of 22385 bp in length
* 41470 41569: gap of 100 bp
* 41570 48545: contig of 6976 bp in length
* 48546 48645: gap of 100 bp
* 48646 55246: contig of 6601 bp in length
* 55247 55346: gap of 100 bp
* 55347 62502: contig of 7156 bp in length
* 62503 62602: gap of 100 bp
* 62603 70135: contig of 7533 bp in length
* 70136 70235: gap of 100 bp
* 70236 78809: contig of 8574 bp in length
* 78810 78909: gap of 100 bp
* 78910 88980: contig of 10071 bp in length
* 88981 89080: gap of 100 bp
* 89081 100598: contig of 11518 bp in length
* 100599 100698: gap of 100 bp
* 100699 115295: contig of 14597 bp in length
* 115296 115395: gap of 100 bp
* 115396 128681: contig of 13266 bp in length
* 128682 128781: gap of 100 bp
* 128782 140301: contig of 11520 bp in length
* 140302 140401: gap of 100 bp
* 140402 150306: contig of 9905 bp in length
* 150307 150406: gap of 100 bp

```



FEATURES

source

104985 104984: contig of 13998 bp in length

105085 105084: gap of unknown length

127380 127379: contig of 22295 bp in length

127480 127479: gap of unknown length

147029 147028: contig of 19549 bp in length

147129 147128: gap of unknown length

168360 168360: contig of 21232 bp in length.

Location/Qualifiers

1. 168360

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="RP11-759J14"

1. 1711

/note="assembly\_name:Contig6"

1812. 6135

/note="assembly\_name:Contig7"

6236. 8687

/note="assembly\_name:Contig8"

clone\_end:SP6

vector\_side:right"

8788. 11647

/note="assembly\_name:Contig9"

11748. 13718

/note="assembly\_name:Contig10"

13819. 18016

/note="assembly\_name:Contig11"

18117. 21570

/note="assembly\_name:Contig12"

21671. 24723

/note="assembly\_name:Contig13"

24824. 32650

/note="assembly\_name:Contig14"

32751. 46491

/note="assembly\_name:Contig15"

46592. 60127

/note="assembly\_name:Contig16"

60228. 74481

/note="assembly\_name:Contig17"

74582. 90886

/note="assembly\_name:Contig18"

90987. 104984

/note="assembly\_name:Contig19"

105085. 127379

/note="assembly\_name:Contig20"

127480. 147028

/note="assembly\_name:Contig21"

147129. 168360

/note="assembly\_name:Contig22"

BASE COUNT 48985 a 34329 c 34643 g 48785 t 1618 others

ORIGIN

Query Match 68.0%; Score 17; DB 71; Length 168360;

Best Local Similarity 100.0%; Pred. No. 9.2;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cgtcagtcgacgtcag 24

|||||

Db 153451 CGTCAGTCGACGTCAG 153435

RESULT 21

AC060833/c

LOCUS AC060833 176379 bp DNA HTG 17-AUG-2000

DEFINITION Homo sapiens chromosome 4 clone RP11-556f3, WORKING DRAFT SEQUENCE, 4 unordered pieces.

ACCESSION AC060833

VERSION AC060833.3 GI:9838272

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 176379)

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished

REFERENCE

2 (bases 1 to 176379)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (20-APR-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108 USA

On Aug 17, 2000 this sequence version replaced g1:8980022.

COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H\_NH0556F03

----- Summary Statistics -----

Sequencing vector: M13; 100%

Sequencing vector: Plasmid; 0%

Chemistry: Dye-primer ET; 100% of reads

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 173750 bases at least Q40

Consensus quality: 174579 bases at least Q30

Consensus quality: 175046 bases at least Q20

Insert size: 174000; agarose-ff

Insert size: 177415; sum-of-contigs

Quality coverage: 6.66 in Q20 bases; agarose-ff

Quality coverage: 6.55 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1419: contig of 1419 bp in length

\* 1420 1519: gap of unknown length

\* 1520 10016: contig of 8497 bp in length

\* 10017 10116: gap of unknown length

\* 10117 79219: contig of 69103 bp in length

\* 79220 79319: gap of unknown length

\* 79320 176379: contig of 97060 bp in length.

FEATURES

source

1. 176379

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="4"

/clone="RP11-556f3"

1. 1419

/note="assembly\_name:Contig5"

1520. 10016

/note="assembly\_name:Contig6"

clone\_end:SP6

vector\_side:right"

10117. 79219

/note="assembly\_name:Contig7"

79320. 176379

/note="assembly\_name:Contig8"

clone\_end:r7

vector\_side:left"

BASE COUNT 49569 a 33984 c 35497 g 57025 t 304 others

ORIGIN

Query Match 68.0%; Score 17; DB 72; Length 176379;

Best Local Similarity 100.0%; Pred. No. 9.1;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 8 cgtcagctcagctcag 24  
 DB 81954 CCGCAGCTCAGCTCAG 81938

RESULT 22  
 AC022781  
 LOCUS  
 DEFINITION Mus musculus chromosome 11 clone RP23-199H17 map 11, WORKING DRAFT  
 AC022781  
 AC022781.5 GI:10305259  
 HTG: HTGS\_PHASE1, HTGS\_DRAFT.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 200729)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Mus musculus chromosome 11, clone RP23-199H17  
 Unpublished  
 2 (bases 1 to 200729)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,  
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,  
 Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
 Deaellano, K., Dewar, K., Domino, M., Doyle, M., Fenesstor, U.,  
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
 Garfield, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
 Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,  
 McPheeters, R., Meldrum, J., Menus, L., Morrow, D., Naylor, J.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., Ollivier, T. M., Peterson, K.,  
 Piere, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
 Tirelli, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
 Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 26, 2000 this sequence version replaced gi:9154653.  
 All repeats were identified using RepeatMasker.  
 Smit, A. F. A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
 Genome Center  
 Center code: W1BR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

Project Information  
 Center project name: L5913  
 Center clone name: 199.H\_17  
 Summary Statistics  
 Sequencing vector: M13: M77815: 100% of reads  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 180464 bases at least Q40  
 Consensus quality: 190837 bases at least Q30  
 Consensus quality: 194945 bases at least Q20  
 Insert size: 206000; agarose-fp  
 Insert size: 197129; sum-of-contigs  
 Quality coverage: 3.9 in Q20 bases; agarose-fp  
 Quality coverage: 4.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
 consists of 37 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will  
 \* be preserved.  
 1 7021: contig of 7021 bp in length  
 \* 7022 7121: gap of 100 bp  
 \* 7122 8310: contig of 1189 bp in length  
 \* 8311 8410: gap of 100 bp  
 \* 8411 9411: contig of 1001 bp in length  
 \* 9412 9511: gap of 100 bp  
 \* 9512 11581: contig of 2070 bp in length  
 \* 11582 11681: gap of 100 bp  
 \* 11682 13342: contig of 1661 bp in length  
 \* 13343 13442: gap of 100 bp  
 \* 13443 16024: contig of 2582 bp in length  
 \* 16025 16124: gap of 100 bp  
 \* 16125 17440: contig of 1316 bp in length  
 \* 17441 17540: gap of 100 bp  
 \* 17541 18922: contig of 1382 bp in length  
 \* 18923 19022: gap of 100 bp  
 \* 19023 21353: contig of 2331 bp in length  
 \* 21354 21453: gap of 100 bp  
 \* 21454 23674: contig of 2221 bp in length  
 \* 23675 23774: gap of 100 bp  
 \* 23775 25868: contig of 2094 bp in length  
 \* 25869 25968: gap of 100 bp  
 \* 25969 27966: contig of 1998 bp in length  
 \* 27967 28066: gap of 100 bp  
 \* 28067 31051: contig of 2985 bp in length  
 \* 31052 31151: gap of 100 bp  
 \* 31152 33236: contig of 2085 bp in length  
 \* 33237 33336: gap of 100 bp  
 \* 33337 36285: contig of 2949 bp in length  
 \* 36286 36385: gap of 100 bp  
 \* 36386 39476: contig of 3091 bp in length  
 \* 39477 39576: gap of 100 bp  
 \* 39577 42584: contig of 3008 bp in length  
 \* 42585 42684: gap of 100 bp  
 \* 42685 46164: contig of 3480 bp in length  
 \* 46165 46264: gap of 100 bp  
 \* 46265 49973: contig of 3709 bp in length  
 \* 49974 50073: gap of 100 bp  
 \* 50074 55389: contig of 5316 bp in length  
 \* 55390 55489: gap of 100 bp  
 \* 55490 59283: contig of 3794 bp in length  
 \* 59284 59383: gap of 100 bp  
 \* 59384 62877: contig of 3494 bp in length  
 \* 62878 62977: gap of 100 bp  
 \* 62978 68182: contig of 5205 bp in length  
 \* 68183 68282: gap of 100 bp  
 \* 68283 73862: contig of 5580 bp in length  
 \* 73863 73962: gap of 100 bp  
 \* 73963 83038: contig of 9076 bp in length  
 \* 83039 83138: gap of 100 bp  
 \* 83139 87822: contig of 4684 bp in length  
 \* 87823 87922: gap of 100 bp  
 \* 87923 92086: contig of 4164 bp in length  
 \* 92087 92186: gap of 100 bp  
 \* 92187 97007: contig of 4821 bp in length  
 \* 97008 97107: gap of 100 bp  
 \* 97108 123018: contig of 25911 bp in length  
 \* 123019 123118: gap of 100 bp  
 \* 123119 129307: contig of 6189 bp in length  
 \* 129308 129407: gap of 100 bp  
 \* 129408 137163: contig of 7756 bp in length  
 \* 137164 137263: gap of 100 bp  
 \* 137264 147510: contig of 10247 bp in length  
 \* 147511 147610: gap of 100 bp  
 \* 147611 155502: contig of 7892 bp in length  
 \* 155503 155602: gap of 100 bp  
 \* 155603 168939: contig of 13337 bp in length  
 \* 168940 169039: gap of 100 bp  
 \* 169040 181336: contig of 12297 bp in length  
 \* 181337 181436: gap of 100 bp  
 \* 181437 195395: contig of 13959 bp in length



gene  
CDS  
10DIGAFKAMQEAVEROVVTYYTDDLTNDKDKTKKIAEVLQAAARLRGVCTIEVH  
VDRVHSMKVIQGDDEVCFSNMFESANRSMAYAKHETSLYRGRLADEROTRLNSLR  
ORITDDPOOVGOA"  
2311..2484  
/gene="XF2410"  
2311..2484  
/note="hypothetical protein: identified by sequence  
similarity: putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF85209.1"  
/db\_xref="GI:9107596"  
/translation="MILSKSARDAVLRRQVDVTGMRQAMHAAEGRDVSPAVALLA  
VALKICCCPLG"  
complement(3181..5514)  
/gene="XF2411"  
complement(3181..5514)  
/note="hypothetical protein: identified by sequence  
similarity: putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF85210.1"  
/db\_xref="GI:9107597"  
/translation="MLKKLIKVNHGILRNGVPTPTPCGPVTLFYAENGRCCTTYA  
NILRAYTHGKRDDEVLOKARTIDTDPPCIELLMNDNTDITKVLHGHWTGPDDL  
IFDAETVDNVSQEVNPAHROKELQFVGAESVKLEREIVRLAKEIADEHNRKIKR  
EORIRASKNMIPADFIRLIPVDAEAKIALRTORAVRNAPALQORHAPQIPHE  
CNDAPFSIIQSTHDIRADAEIYQAHFORHOOPGIEAMVMOGRFENEGCPFG  
OEVTATLIKAYETETNTAFKAPKNVATLGSVATREPCLEKTIADRIQSNACVA  
AMKQDLSPSPDNPRATATPTPLVALALASKORHPDCTTQDADNRYALLE  
ELRQMAHYANAWEINRITNFRKTI AENTPALDLSITLIGIVRQSTEVVQATLE  
YQAAKAKKEQLEEREKKNMRALDAGLPDLSEVASKITNOFLRDPGAFSIFKLOQSM  
QCGTRASVYLQLRGKEVALGRTRSDPGFHNVSIEDKRTLALAFIARLYATPDL  
VGKSVLDDPWCSEFDMTRNMTESIALVNOGVVVLSDAYETLRLDRLAGPPL  
NKSVNVHIIKRTKNDISOIYSDADLTICQSPMLRYAOVYTFVSGTGTLOEVS  
ALRLVSGFLKHPRAPLLRQDLSLQGMISIRKATPDSPLVLAQPIVDTLEKLNARA  
VOSHDDSKSLSPMNDGOLROYARMALELYGGSLPH"  
complement(5615..5800)  
/gene="XF2412"  
complement(5615..5800)  
/note="hypothetical protein: identified by sequence  
similarity: putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF85211.1"  
/db\_xref="GI:9107598"  
/translation="MIFGATQTNADRRRSGTRHACTGMAERSRLPPDPVRIHAHL  
KPHKVISSIPITGTL"  
6084..6512  
/gene="XF2413"  
6084..6512  
/note="hypothetical protein: identified by sequence  
similarity: putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF85212.1"  
/db\_xref="GI:9107599"  
/translation="MWGGDVPTADFRRAAGACEPSIKNTALSGVHLLDLCAGRLM  
PGCCIGGNGPMKGVKSVKSVLGVLRCGSCQGVAAARFAPRTFGKYCMSALML  
YRWCSERHLCSTKMEGATSGDAGPRRGRLGKGRG"  
complement(6509..8314)  
/gene="XF2414"  
complement(6509..8314)  
/note="hypothetical protein: identified by sequence

similarity: putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF85213.1"  
/db\_xref="GI:9107600"  
/translation="MPPKRRHOVDAADTKENLTVTTEHCFIGVCFPIPDGSGQREFG  
FSEFLGALMLAMTIAVRYRFRIOVAPLPLRMSFYVVAAGIGLITLTPMBAEO  
MIVPRGILLTPAVMOALIGAFLLITLNMAGIFIRPPYGOVNAARPOQTIYRIILN  
GSPTELAVIADLVHSAELVQAHATNRKSIQNGERTICEEEDDLPVNRDANALL  
LIAEKRLCRAIVAESPTAFALFOALANRKYGIPIRYAQNIFHEALANKDSFLYHE  
TNGCALGLIGDHKFTQAMFANEWEITGTLFNHRDARNRMDEPOMIATYCSVLMT  
IRSYAKQGRGYSEVLQAMGYTASAASDLNPESSAASDVLRLALVIOFIKDA  
VRILDEGVGEHCICIRAGYDSSRAIFCDRIANMIFVYISLAEVRLPOLMIOH  
SVNSNLPHGAEFRPAGOVAFKRLRLITNOADRMQFNVERAKILRECLAVLAV  
IRGNDYKSDRRFHRHRYLHRAVLAMTKHVMHSHHPRLAESCLGNHTTYDDEHRLA  
YTHPAKPGIKVSYFALNPASPPPEVVAASAVPNA"  
complement(8365..8440)  
/gene="trna-thr-2"  
/note="found by tRNAscan"  
/product="trna-thr"  
complement(8365..8440)  
/gene="trna-thr-2"  
complement(8457..8648)  
/gene="XF2415"  
complement(8457..8648)  
/note="hypothetical protein: identified by sequence  
similarity: putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF85214.1"  
/db\_xref="GI:9107601"  
/translation="MPVPAVDPPOPPIKSHLCATVAHLQSTSPDALLIMRRAP  
EAHTKOSALSRRPSTAM"  
complement(8921..9871)  
/gene="XF2416"  
complement(8921..9871)  
/note="similar to GI:5915671 (percent identity: 64 %/query  
alignment coverage: 99.4 %/subject alignment coverage:  
99.4 %); identified by sequence similarity: putative; ORF  
located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="drug tolerance protein"  
/protein\_id="AAF85215.1"  
/db\_xref="GI:9107602"  
/translation="MKVILLANPRGAGVDRALIEYKRTIDMIGTPIYVRRHVHNR  
VVDLIKORCAIYVEELHVOYPDGATVYFSAHGYSQAVRRQARGLKVPDAPPLVTKV  
HLUDVANHCRGRDMILIGHAGHREVEGTGQWDQEGTGRITLVEIDIVATLHVAOP  
HHLAVYQPTTLVSDOTRNTIDALROEPTIQCKNNDICATQONRODAVRELARCDL  
VLVAGSPNSNSNRLSIELAORBGVASYLIDSAEIDPAVYIDKHNIIGVYAGASAPVL  
VDCVLARLYELGATVSSEHSGKPESWFALPALRLQLVD"

Query Match 64.0%; Score 16; DB 1; Length 10536;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cytcaggtgcagctca 23  
|||||  
Db 1995 CGTCAGGTGCAGGTCA 2010

RESULT 24  
AC004508  
LOCUS  
DEFINITION Homo sapiens chromosome 21, p1 clone LBL43 (UBNL H09), complete  
sequence.  
AC004508 L35684 L35685 L35686 L35687 AC000939 AC000938 AC000940  
AC000937 LB1573 L35689 L35690 L35691 L35692 L35693 L35694 L35695  
L35696 L35697 L35698 L35699 LA2077 L35700 L35701





-----  
Quality coverage: 4.4 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1 1068: contig of 1068 bp in length
* 1069 1168: gap of 100 bp
* 1169 2717: contig of 1549 bp in length
* 2718 2817: gap of 100 bp
* 2818 6040: contig of 3223 bp in length
* 6041 6140: gap of 100 bp
* 6141 10201: contig of 4061 bp in length
* 10202 10301: gap of 100 bp
* 10302 14615: contig of 4314 bp in length
* 14616 14715: gap of 100 bp
* 14716 21086: contig of 6371 bp in length
* 21087 21186: gap of 100 bp
* 21187 25767: contig of 4581 bp in length
* 25768 25867: gap of 100 bp
* 25868 32500: contig of 6633 bp in length
* 32501 32600: gap of 100 bp
* 32601 38651: contig of 6051 bp in length
* 38652 38751: gap of 100 bp
* 38752 50922: contig of 12171 bp in length
* 50923 51022: gap of 100 bp
* 51023 71014: contig of 19992 bp in length
* 71015 71114: gap of 100 bp
* 71115 93563: contig of 22449 bp in length
* 93564 93663: gap of 100 bp
* 93664 123080: contig of 29417 bp in length
* 123081 123180: gap of 100 bp
* 123181 155628: contig of 32448 bp in length.
Location/Qualifiers
1. 155628

```

## FEATURES

```

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-357J12"
/clone_1ib="RPC1-11 Human Male BAC"
1. 1068
misc_feature
/feature="assembly_fragment"
1169..2717
/feature="assembly_fragment"
2818..6040
misc_feature
/feature="assembly_fragment"
6141..10201
misc_feature
/feature="assembly_fragment"
10302..14615
misc_feature
/feature="assembly_fragment"
14716..21086
misc_feature
/feature="assembly_fragment"
21187..25767
misc_feature
/feature="assembly_fragment"
25868..32500
misc_feature
/feature="assembly_fragment"
32601..38651
misc_feature
/feature="assembly_fragment"
38752..50922
misc_feature
/feature="assembly_fragment"
51023..71014
misc_feature
/feature="assembly_fragment"
71115..93563
misc_feature
/feature="assembly_fragment"
93664..123080
misc_feature
/feature="assembly_fragment"
123181..155628
/feature="assembly_fragment"

```

```

BASE COUNT      37665 a 38986 c 40608 g 37067 t 1302 others
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 71; Length 155628;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      9 gtaggtgcaggtcag 24
Db 95800 GTCAGGTGCAGGTAC 95785

```

```

RESULT 27
AC005412/c
LOCUS
DEFINITION
AC005412
ACCESSION
AC005412.6 GI:11128436
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
1 (bases 1 to 183700)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone hRPK.22.N.12
Unpublished
2 (bases 1 to 183700)
Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E.,
Devon,K., Dewar,K., Donelan,L., Ferreira,P., Fitzhugh,W.,
Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraldery,K., Grant,G.,
Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C.,
Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J.,
Nafit,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Testfay,S., Tichovolsky,N., Torrella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (12-AUG-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 183700)

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

Direct Submission
Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 183700)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckery,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Locke,K., Macdonald,P., Marquis,N.,
Lehoczy,J., Liu,C., Locke,K., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testfay,S., Titrrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

```

Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heatord, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, D., Lenoczky, J., Lieu, C., Locke, K., Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (09-NOV-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141 USA  
 On Nov 9, 2000 this sequence version replaced gi:5931398.  
 All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

## FEATURES

## source

1. 183700  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="hRPK.22.N.12"  
 /clone\_lib="RPCL-11 Human Male BAC"  
 1. 70  
 /rpt\_family="MER41D"  
 784. 1084  
 /rpt\_family="AluY"  
 1320. 1637  
 /rpt\_family="AluJo"  
 complement(1664. 1959)  
 /rpt\_family="AluSx"  
 1961. 2074  
 /rpt\_family="L2"  
 2283. 2355  
 /rpt\_family="C-rich"  
 complement(2369. 2774)  
 /rpt\_family="LIM4"  
 2775. 3066  
 /rpt\_family="AluSg"  
 complement(3067. 3077)  
 /rpt\_family="LIM4"  
 3549. 3814  
 /rpt\_family="AluSg"  
 complement(3867. 3949)  
 /rpt\_family="AluJo"  
 complement(3950. 4247)  
 /rpt\_family="AluJb"  
 complement(4248. 4361)  
 /rpt\_family="AluJo"  
 complement(4362. 4490)  
 /rpt\_family="LIM4"  
 complement(4493. 4589)  
 /rpt\_family="LIM4"  
 4590. 4888  
 /rpt\_family="AluSx"  
 complement(4889. 5305)  
 /rpt\_family="LIM4"  
 complement(5328. 5623)  
 /rpt\_family="AluSx"  
 complement(5876. 6164)  
 /rpt\_family="AluSg"  
 6196. 6509  
 /rpt\_family="AluSg"  
 6525. 6580  
 /rpt\_family="AT-rich"  
 complement(6671. 6784)  
 /rpt\_family="L2"  
 complement(6986. 7137)  
 /rpt\_family="AluSg"  
 complement(7138. 7447)  
 /rpt\_family="AluSx"

repeat\_region complement(7448. 7582)  
 /rpt\_family="AluSg"  
 repeat\_region 7808. 8106  
 /rpt\_family="AluJb"  
 unsure 7825. 8084  
 /note="Single-stranded terminator coverage."  
 repeat\_region 8121. 8476  
 /rpt\_family="LIMB5"  
 repeat\_region 8496. 8677  
 /rpt\_family="LIMB5"  
 /rpt\_family="LIMB5"  
 8503. 8791  
 /note="Single-stranded terminator coverage."  
 repeat\_region 8689. 8870  
 /rpt\_family="LIMB5"  
 /rpt\_family="LIMB5"  
 8736  
 /note="Probably T, possibly C."  
 repeat\_region 8877. 9144  
 /rpt\_family="LIMB5"  
 repeat\_region 9147. 9210  
 /rpt\_family="LIMB5"  
 repeat\_region 9216. 9395  
 /rpt\_family="LIMB5"  
 /note="Single-stranded terminator coverage."  
 repeat\_region 9458. 9530  
 /rpt\_family="LIMB5"  
 repeat\_region 9536. 9848  
 /rpt\_family="LIMB5"  
 repeat\_region 9878. 10115  
 /rpt\_family="LIMB5"  
 repeat\_region 10128. 10357  
 /rpt\_family="LIMB5"  
 repeat\_region 10432. 10458  
 /rpt\_family="LIMB5"  
 repeat\_region 10612. 10908  
 /rpt\_family="LIMB5"  
 repeat\_region 10913. 11210  
 /rpt\_family="LIMB5"  
 repeat\_region 11216. 11512  
 /rpt\_family="LIMB5"  
 repeat\_region 11529. 11553  
 /rpt\_family="LIMB5"  
 repeat\_region 11647. 11759  
 /rpt\_family="LIMB5"  
 repeat\_region 11816. 12159  
 /rpt\_family="LIMB5"  
 repeat\_region 13008. 13306  
 /rpt\_family="LIMB5"  
 repeat\_region 13355. 13490  
 /rpt\_family="LIMB5"  
 repeat\_region 13491. 13789  
 /rpt\_family="LIMB5"  
 repeat\_region 13790. 13956  
 /rpt\_family="LIMB5"  
 repeat\_region 14078. 14206  
 /rpt\_family="LIMB5"  
 repeat\_region 14220. 14276  
 /rpt\_family="LIMB5"  
 repeat\_region 14277. 14309  
 /rpt\_family="LIMB5"  
 repeat\_region 14310. 14594  
 /rpt\_family="LIMB5"  
 repeat\_region 14604. 14912  
 /rpt\_family="LIMB5"  
 repeat\_region 14913. 15045  
 /rpt\_family="LIMB5"  
 repeat\_region 15061. 15157  
 /rpt\_family="LIMB5"  
 repeat\_region 15517. 15637  
 /rpt\_family="LIMB5"

Query Match

64.0%; Score 16; DB 86; Length 183700;

Best Local Similarity 100.0%; Pred. No. 32;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 gtcaggtgcagtcac 24  
|||||  
Db 177274 gtcaggtgcagtcac 177259

RESULT 28  
LOCUS AE001862/c  
DEFINITION Deinococcus radiodurans RI section 1 of 2 of the complete  
ACCESSION AE001862 AE001825  
VERSION AE001862.1 GI:6460468  
KEYWORDS  
SOURCE Deinococcus radiodurans.  
ORGANISM Deinococcus radiodurans group; Deinococcales; Deinococcus.  
REFERENCE 1 (bases 1 to 213732)  
AUTHORS White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D.,  
Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L.,  
Moffat, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
Vamathevan, J. J., Lam, P., McDonald, L., Utecherack, T., Zaleski, C.,  
Makarov, K. S., Aravind, L., Daly, M. J., Fraser, C. M. et al.  
Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans RI  
Science 286 (5444), 1571-1577 (1999)

TITLE 2 (bases 1 to 213732)  
JOURNAL White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D.,  
MEDLINE Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L.,  
REFERENCE Moffat, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
AUTHORS Vamathevan, J. J., Lam, P., McDonald, L., Utecherack, T., Zaleski, C.,  
Makarov, K. S., Aravind, L., Daly, M. J., Minton, K. W.,  
Fleischmann, R. D., Ketchum, K. A., Nelson, K. E., Salzberg, S.,  
Smith, H. O., Venter, J. C. and Fraser, C. M.  
Direct Submission  
Submitted (02-NOV-1999) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
location/Qualifiers  
1. 213732  
/organism="Deinococcus radiodurans"  
/strain="RI"  
/db\_xref="taxon:1299"  
/chromosome="2"  
653. .1435  
/gene="DRA0001"  
653. .1435  
/note="similar to GB:D26185 SP:P37522 PID:467381  
PID:580906 GB:AL009126 percent identity: 54.96; identified  
by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="chromosome partitioning ATPase, putative, ParA  
family"  
/protein\_id="AAFI2301.1"  
/db\_xref="GI:6460593"  
/translation="MVSAYTTLVFNHAGAGTSLTLNGLYELARGLRLLVLLDIP  
QANLITGLIGSTREMTYVPAVDQPLSPYKAGLDVIPAHSVLAIAEQGMGRV  
GAQGLRRLAEVSDVALIDSPSLGOLATLAAADOMIVPPTKQGLDALPG  
LOGALTEYREVRDLTVALYVPTFYDARRRHOEVLADKAHLSPILARVPDREAVWL  
DSTAAGAPSEVAPGTPVHADVGRRLADIAALICGAVPGENA"  
1432. .2313  
/gene="DRA0002"  
1432. .2313  
/gene="DRA0002"  
/note="similar to GB:I34077 GB:X85964 PID:407375  
PID:757761 PID:1163135 percent identity: 57.34; identified  
by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1

/product="chromosome partitioning protein, ParB family"  
/protein\_id="AAFI2301.1"  
/db\_xref="GI:6460594"  
/translation="MTRRRPERRRDLGLGTFPVDSQANDIRALPVNELKVGSTOP  
RRSPDLERLSEIASTRAHGVLOPLVRSVQGYEIVAGERRRAQLAGLEAVPV  
ROLSNEQARAALITENLRDNLVIDEVSKELELLATGLLEERFAKRLMOLRAVP  
GDEHEQIDQYFRSMGETWTFPAKRLRLIMPPVYELARAGIPILIGSVVASPER  
QAEELTLAONGASRSOLLQALQTPSOTSVAVTPHFAKVLVSSKRFSLGLDTPREALDR  
WLAMPERRVQAIDEQS"  
complement(2418. .3881)  
/gene="DRA0003"  
complement(2418. .3881)  
/note="similar to GB:M88334 SP:P25526 PID:147901 GB:U00096  
PID:11789015 percent identity: 69.60; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="succinate-semialdehyde dehydrogenase"  
/protein\_id="AAFI2294.1"  
/db\_xref="GI:6460588"  
/translation="MTTDLRTYSSVTSQAYFDGEWRNAPNPFVHPGCEVIGE  
VADCTPDARQAIIDAEVALREMOVNPYERGIILRMHDLMEHREBELAQMLTMBG  
KPISETGEVYASFIEMCAEEGRJAGERILRFPKRGITISEPVGYVAVTWN  
PPAMITRKAPALAAAGCVMILKPAELSPMTALYLETMLKGGAPATFOVLDLDER  
ALTPPNDSVRRLTTCSTEVGRLLYQOAGITKRSLELGGHAPLTFDDADLR  
AASEVVASKEFNSSGCTCCTNRVYVQGVAAEFTRLTETALQGLDPPDEYQGP  
VVEQAGLDKRVQVQDALTKGAQVTTGGVSSGFLPQVPLVAVAPSLRLREETGP  
VAPVTFIDTEEEGLRLANDSEYGLAAVAVTRDGRARLAEGLYGVIGINDGLPSSA  
APHYFGKMSGVGRGHWGLLEYLETFVSGLS"  
4024. .4668  
/gene="DRA0004"  
4024. .4668  
/note="similar to GP:2924378 percent identity: 55.29;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAFI2299.1"  
/db\_xref="GI:6460593"  
/translation="MNSLSIAAVLLMAGSTRMGOPKALLPLGGMPLCHAAOTIVE  
AGYGEVAVLPPGETGDVRAALQGLPLRWYTNQPPQGLSSQAAEGELMLESAP  
AVVETLADMLPVRDTHQALGTVGAGRPAAITIRXGEVAPATPLLRDLRLPRHEP  
PSDHGPRTLIRELRELBELKXIDRPAAEILIDTPPAAQAOQLFAEYVPGAQDGP"  
complement(4719. .5897)  
/gene="DRA0005"  
complement(4719. .5897)  
/note="similar to GB:I33464 SP:P47734 PID:496118 percent  
identity: 74.23; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="alcohol dehydrogenase, zinc-containing"  
/protein\_id="AAFI2293.1"  
/db\_xref="GI:6460587"  
/translation="MAKATVQGVNKKVGEVPPDLLPTDAIVKITSTACSDLHL  
LDGYTPMSEGDIIIGHFEMGEVVEVSEVRLLKGVDRVYVPEFAGVCPCKRGFS  
ACDNSNPVHRRAEALXGATSGCALPGYSHMGGVAGGAAOAVRVFPADVPKTESL  
KDEQVLFITDIFPGYQAAEOCNIVPGRDVAVYAGAPVGFARSRQMGAAHVITY  
DVQPERLMAAASACQITINERKENVLSLRATGRCGDVHIDAIVGAHSGHGSIM  
DKVQCNLKLTFDRITALEMAITLSCAGTGVSMGPVGLVDKMPGIAFAAGLIRMG  
QTHRHRIPIILIGRIEAGQIDPSFVITHRASLDOAPELVKTFRDKQCCIKVYLPWA  
complement(5894. .6700)  
/gene="DRA0006"  
complement(5894. .6700)  
/note="similar to PID:1653488 percent identity: 48.46;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1



```

RESULT 29      HSMX1B      313064 bp      DNA      PRT      23-SEP-2000
LOCUS         HSMX1B
DEFINITION    Homo sapiens chromosome 21 from 5 PACs and 5 Cosmids map
               21q22.2,D21S349-MX1; segment 2/2, complete sequence.
ACCESSION     AL442167
VERSION       AL442167.1   GI:10303260
KEYWORDS      HTG.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 313064)
AUTHORS       Ramser,J., Francis,F., Beck,A., Hennig,S., Klages,S., Borzym,K.,
               Langer,I., Steffens,C., Hildmann,T., Dagand,E., Yaspo,M.,
               Reinhardt,R. and Lehrach,H.
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 313064)
AUTHORS       MPIMG.
TITLE         Direct Submussion
JOURNAL       Submitted (01-OCT-1998) MPIMG, Abt. Lehrach, Max Planck Institut
               fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195, Germany
COMMENT       Clones received from Resource Centre of the Human Genome Project at
               the Max-Planck-Institut for Molecular Genetics.
               292889. .294336 sequence from clone KB447A5, accession no. AP001609
               (DBJ), sequenced at KEIO university, Tokyo, Japan.
               Bases 1. .50 overlap with AL442166.
FEATURES             location/Qualifiers
                     1..313064
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="21"
                        /map="21q22.2.D21S349-MX1"
                        /clone="PAC RPCI-1.247E2"
                        /clone="PAC RPCI-1.146B4"
                        /clone="PAC RPCI-1.141D16"
                        /clone="PAC RPCI-1.269A14"
                        /clone="PAC RPCI-1.265B9"
                        /clone="cosmid LNLc116.44C5"
                        /clone="cosmid LNLc116.16H18"
                        /clone="cosmid LNLc116.14C10"
                        /clone="cosmid LNLc116.25D2"
                        /clone="cosmid LNLc116.87D5"
                        /clone="lib=LI2INC02, Chromosome 21 specific cosmid
                        library, Lawrence Livermore National Laboratory (LNL),
                        creator: Pieter de Jong; PAC: RPCI1,3-5, Roswell Park
                        Cancer Institute, creator: Pieter de Jong, P.Ioannou"
                        292889. .294336
                        (/note="sequence from clone KB447A5, accession no. AP001609
                        (DBJ)", sequenced at KEIO university, Tokyo, Japan."
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="21"
SOURCE         ORIGIN      83213 a 71272 c 73589 g 84990 t
Query Match          64.0%; Score 16; DB 93; Length 313064;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Oy 9 gtcaggtgcaggtcac 24
Db 182813 CTCAGGTGCAGTCAg 182828

```





QY 10 tcaggtcaggtcag 24  
|||||  
Db 408 TCAGGTGCAAGTCAG 422

RESULT 33  
TCU57884  
LOCUS Thanatephorus cucumeris strain 2tr144 5.8S ribosomal RNA gene and  
DEFINITION internal transcribed spacers 1 and 2, complete sequence.  
ACCESSION U57884  
VERSION 1  
KEYWORDS GI:1914814  
SOURCE  
ORGANISM Thanatephorus cucumeris.  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Hymenomycetidae;  
Ceratobasidiales; Ceratobasidiaceae; Thanatephorus.  
REFERENCE 1 (bases 1 to 728)  
AUTHORS Salazar, O., Schneider, H., Keijer, J. and Rubio, V.  
TITLE Phylogenetic relations in Rhizoctonia solani AG2 using ITS  
sequences  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 728)  
AUTHORS Rubio, V.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-1996) Victor Rubio, Biotechnology Microbiana,  
Centro Nacional de Biotecnologia (CSIC-UAM), Campus  
Cantoblanco-UAM, Madrid, 28049, Spain  
FEATURES  
source  
1. 728  
/organism="Thanatephorus cucumeris"  
/strain="2tr144"  
/specific\_host="Tulipa sp."  
/db\_xref="taxon:118239"  
/tissue\_type="mycelia"  
/note="Isolated from tulip, from The Netherlands"  
1. 294  
/note="Internal transcribed spacer 1; ITS1"  
295..449  
/product="5.8S ribosomal RNA"  
450..728  
/note="Internal transcribed spacer 2; ITS2"  
BASE COUNT 196 a 153 c 145 g 234 t  
ORIGIN

Query Match 60.0%; Score 15; DB 15; Length 728;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 cctcgtcaggtcag 19  
|||||  
Db 20 CCTGTCAGGTGAG 34

RESULT 34  
XLHNF4MR 1552 bp mRNA VRT 07-MAR-1996  
LOCUS X.laevis HNF4 mRNA encoding hepatocyte nuclear factor 4 (HNF4).  
DEFINITION 237526  
VERSION 1  
KEYWORDS GI:1150689  
hepatocyte nuclear factor 4.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Xenopus.  
REFERENCE 1 (bases 1 to 1552)  
AUTHORS Holtsch, B., Strandmann, E.P., Zapp, D., Lorenz, P. and Ryffel, G.U.  
TITLE Transcriptional hierarchy in Xenopus embryogenesis: HNF4, a maternal  
factor involved in the developmental activation of the gene  
encoding the tissue specific transcription factor HNF1 alpha (LFB1)  
JOURNAL Mech. Dev. 54 (1), 45-57 (1996)

MEDLINE 96404127  
REFERENCE 2 (bases 1 to 1552)  
AUTHORS Ryffel, G.U.  
TITLE Direct Submission  
JOURNAL Submitted (16-SEP-1994) Ryffel, G. U., Universitaetsklinikum,  
Institut fuer Zellbiologie, Hufelandstrasse 55, Essen, Germany,  
D-45122

FEATURES  
source  
1. 1552  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/sex="Male"  
/dev\_stage="adult"  
/tissue\_type="liver"  
/clone\_lib="liver cDNA"  
93..1460  
/standard\_name="HNF4"  
/function="transcription factor"  
/citation=[1]  
/codon\_start=1  
/evidence=experimental  
/product="hepatocyte nuclear factor 4 (HNF4)"  
/protein\_id="CA85763.1"  
/db\_xref="GI:1150690"  
/db\_xref="SPRMBL:Q91766"  
/translation="MDMADYTEALDPAYTTLLEFNMOVLSIGTDTSTSDVTSLSASNS  
IGINSLAICGDRATGKHYGASCDGCKGFRRSVRNHMYSCFSSQCVVDKRMQ  
CRYRLKCKFRAGMKKEAVONERDRISTRSSYEDSLPSINVLIOAEVLSQITSSV  
GLVNTDIDGKIKIACIDVDSMKQOOLLVEMAKYIPAPCELPDDOVALLRAGGH  
LLGNTDKSMFKDILLGNDRILPRNCPLEVGAVRARIIDELVYLPFOELQIDNXY  
ACIKATLFPDPDAGLSPPKIKRMROYOVSYEDYINDQYDSRGFRGELLILPPL  
OSTWOMTEQIQYKRLGMAKINDLIDMELGGGSAKSHTHHRLHPLVODHIAVW  
IVANNVLPOLHNGOMSTPETPQSPAGSAEOYKIVHGTIASINKOPISIPSTTT  
KOEAM"

BASE COUNT 453 a 352 c 378 g 369 t  
ORIGIN

Query Match 60.0%; Score 15; DB 8; Length 1552;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caggtcaggtcag 25  
|||||  
Db 1002 CAGGTGCAAGTCAGC 1016

RESULT 35  
AB056427/c  
LOCUS AB056427  
DEFINITION Macaca fascicularis brain cDNA, clone:Qf1A-15457.  
ACCESSION AB056427  
VERSION 1  
KEYWORDS GI:13358943  
f1s (full insert sequence).  
SOURCE Macaca fascicularis adult male frontal lobe left cDNA to mRNA,  
clone\_11b:macaque brain cDNA library Qf1A clone:Qf1A-15457.  
ORGANISM Macaca fascicularis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
Cercopitheciinae; Macaca.  
REFERENCE 1 (sites)  
AUTHORS Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,  
Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.  
TITLE Isolation of full-length cDNA clones from macaque brain cDNA  
libraries  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1727)  
AUTHORS Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.  
TITLE Direct Submission  
JOURNAL Submitted (24-FEB-2001) Katsuyuki Hashimoto, National Institute of  
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama  
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
(E-mail: khashim@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/,

## COMMENT

Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181

Lab host: TOPI0  
PME18S-FL3 (ACC. No. AB009664)

R. Site1: DraIII (CACTGTGTG)  
R. Site2: DraIII (CACCATGTG)

Description: 1st strand cDNA was primed with an oligo(dT) primer using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of PME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing  
(5' end primer [CGACTGCTCAAAAGCTCGG];  
3' end primer [CGACTGCTCAAGTCAACGCA])  
Location/Qualifiers

## FEATURES

## source

1..1727  
/organism="Macaca fascicularis"  
/db\_xref="taxon:9541"  
/clone="Ofla-15457"  
/sex="male"  
/tissue\_type="frontal lobe left"  
/clone\_lib="macaque brain cDNA library Ofla"  
/dev\_stage="adult"  
1029..1361  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BA033085.1"  
/db\_xref="GI:13358944"  
/translation="MGPSIPAKTREVLSHLSYNNMALOGIEFVAOLKSLVLTGL  
IDHLTVEQAVLRLSEERYIQOKWNIEMAHDYELQELRANTAGTLEFHCSSTT  
VHKLINE"

## CDS

BASE COUNT 468 a 410 c 452 g 397 t  
ORIGIN

Query Match 60.0%; Score 15; DB 85; Length 1727;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 gtcaagtcagatga 23  
|||||  
Db 1177 GTCAGGTGCGAGTCA 1163

## RESULT 36

## LOCUS

RICCYTALD 1801 bp DNA PLN 14-APR-2000  
DEFINITION Oryza sativa gene for cytoplasmic aldolase, complete cds,  
clone:Aldc-a.

ACCESSION D13512.1 GI:218156  
VERSION D13512.1

KEYWORDS cytoplasmic aldolase.

SOURCE Oryza sativa DNA, clone:Aldc-a.

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 1801)

1 Tsutsuni,K., Kagaya,Y., Hidaka,S., Suzuki,J., Tokairin,Y.,  
Hirai,T., Hu,D.L., Ishikawa,K. and Ejiri,S.  
Structural analysis of the chloroplastic and cytoplasmic  
aldolase encoding genes implicated the occurrence of multiple loci  
in rice

## TITLE

Gene 141 (2), 215-220 (1994)

JOURNAL MEDLINE 94215906

REFERENCE 2 (bases 1 to 1801)

AUTHORS Tsuchiya,T.  
TITLE Submitted Submission  
JOURNAL Submitted (28-OCT-1992) to the DDBJ/EMBL/Genbank databases. Tohru  
Tsuchiya, Iwate University, Department of Agriculture, 3-18-8 Ueda,  
Morioka, Iwate 020, Japan (Tel:0196-23-5171(ex.2776)).

## COMMENT

Fax:0196-25-7667)  
Submitted (28-OCT-1992) to DDBJ by:  
Tohru Tsuchiya  
Institute for Cell Biology and Genetics  
Faculty of Agriculture  
Iwate University  
3-18-8 Ueda  
Morioka, Iwate 020  
Japan  
Phone: 0196-23-5171 x2776  
Fax: 0196-24-5084

## FEATURES

## source

Location/Qualifiers  
1..1801  
/organism="Oryza sativa"  
/db\_xref="taxon:4530"  
/clone="Aldc-a"  
TATA.signal 41..49  
prim\_transcript 80..1801  
exon 80..179  
/number=1  
join(152..179,615..1663)  
CDS join(152..179,615..1663)  
/codon\_start=1  
/product="cytoplasmic aldolase"  
/protein\_id="BA02729.1"  
/db\_xref="GI:218157"

/translation="MSAYCGKYKDELKNAAYIGTPGKGLADESTGTIGKRPASIN  
VENEDNRAARELIFCTPGALQYISGILFDEFLYOKTKGRFPVILKEAGALPFI  
KVDGTIEVAGTDEKETTQGHDDGOCARKEAGARPAKRAVYKIGPNPSCGLAD  
LNAOLACAYATICQENGLVPIVEPEIIVDGHDDRCAYVEVYLAACYALNHHYL  
LEGTILKNNMTGPSDAKVAPEVIAETVYTLDRTPPAVPAIVPLSGQSEEARL  
NLNANRKSAPKPSLSFSGRAIQOSTLTKAWGRTVEKARAFIVRCKANSEARL  
GTYGDVAVLGGAESLHVKKYK"  
180..614  
/number=1  
exon 615..1801  
/number=2

BASE COUNT 418 a 521 c 439 g 423 t  
ORIGIN

Query Match 60.0%; Score 15; DB 14; Length 1801;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctctgtcagatgta 18  
|||||  
Db 1551 TCCTCGTCAGGTGCA 1565

## RESULT 37

## LOCUS

HAMHPP 2081 bp mRNA ROD 27-APR-1993  
DEFINITION Mesocricetus auratus pregnancy protein (HPP), complete cds.

ACCESSION M96650.1 GI:191387  
VERSION M96650.1

KEYWORDS pregnancy protein; pregnancy protein 60kDa.  
Mesocricetus auratus (strain Waterhouse) (library: lambda gt 11)  
female adult, 14 days pregnant liver (maternal) cDNA to mRNA.

SOURCE Mesocricetus auratus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Mesocricetus.

REFERENCE 1 (bases 1 to 2081)

1 Cloning and sequencing of a new pregnancy marker protein from  
hamster liver

AUTHORS Hamster liver  
TITLE Tthesis (1992) Biological Sciences, The Wichita State University  
JOURNAL Location/Qualifiers

1..2081  
/organism="Mesocricetus auratus"  
/strain="Waterhouse"  
/db\_xref="taxon:10036"  
/dev\_stage="adult, 14 days pregnant"

## FEATURES

## source





```

SGRRPATLTYTNPFKQAEBSLEDELFKFLADMRSSLLRLRPVTAQTKIDISAP
ENRHFCSPELLHIKTEYEDPRGRTEKLEFPAEYVAPHTSYRNLIVYPHSINSS
RQGSVRNLAVAVOIMTEGDEPSQALPYIFGSSQSEFREAFTPVYHNKSPFEYEFK
LHLPACVTEHNLHLEFHYVSCOPRGTALETPGFTWIPLQHRRLRMLRPFELPVSV
DOPPSYSLTPDVALPGMRWDGKGVSELTAAVSVHPODYLDRKPFTHVLEE
GAPFERKOTVLSGNANVEOELRASIALRLASPEPIYVAFSHHIDKLVIRIPPLIS
GOIYNLGRGAFEMAHYVSLVHRSLEAODARHCQOLAAVYHARLRPTEPSLEPDG
APPVYQATLARGSSGRPASLYLARSKSISSNPDVAVAGSVDEVSRLIASKLLEH
ELALQWVSSAAVEALLOHAFEPFOLMVKSMALHLLGQRLDTPKRLPRGRFELDI
TALVSGVLEVTIRVHKVDELAEHLNANSLAFPLSDLSLVDGEFVSIVAHYKQVAT
RLOSPPNALTLRMEFTRILCSHEHYVTLNPPCLSPSPSPSVSTTSOSSSTF
SSQAPDKPTSMELSGPFOHEFLAGLITLALAEPEAGAPILHKKALISAVSL
LCGHDDPRYAEATVAKARVAEILYPLISTARDTLPLRLEPAEGORSLRSLASLDDT
EGEGDLACTINPSVMAIAGGPIAPSGRASIISGPPAPASRGCALSERSRILLACVL
WVKLNTPEALLQWADLTLPQLGRLLDLTLCLAAFEYKKGKAFERINSITFKSID
MKARLEAILGTIGARQEWYRKRSPENPNPNVKSVTIMKQTSRVDKXDEM
EHLALVGNLALVSKPELLFEEDTELCADCLRLRHGSRISTIRTHASALYLMR
ONEEIGNFARVMQVYMSLSLVGTTQNSFEHLRSLKLTITVAEPMGLDSTFA
EYODLMFNHMLITDLYVKMKEHQEDPEMLDLMYRIARQYQSSPOLRLTLWNNAKX
HAELGNHAEAAQGVNHA"
BASE COUNT      941 a      1605 c      1425 g      915 t
ORIGIN

Query Match      60.0%; Score 15; DB 85; Length 4886;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3      atccctgcaggtgc 17
          |||||
Db      735      ATCCTCGTACAGTGC 721

RESULT 42
LOCUS      AF250776          5526 bp      DNA      BCT      31-JAN-2001
DEFINITION      Uncultured bacterium pCosHE2 hypotheical 17.1 kDa protein in
                  modc-bioc intergenic region, DAPA-aminotransferase BioC (bioc),
                  biotin synthase BioB (biob), KAPA synthetase BioF (biof), and
                  biotin biosynthesis protein BioC (bioc) genes, complete cds; and
                  decthiobiotin synthetase BioD (bioid) gene, partial cds.
ACCESSION      AF250776.1      GI:12620124
VERSION      AF250776
KEYWORDS
SOURCE      uncultured bacterium pCosHE2.
ORGANISM      Bacteria; environmental samples.
REFERENCE      1 (bases 1 to 5526)
AUTHORS      Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streit, W.R.
TITLE      Direct cloning from enrichment cultures, a reliable strategy for
          isolation of complete operons and genes from microbial consortia
          Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
JOURNAL      JOURNAL
MEDLINE      20575196
PUBMED      11133432
2 (bases 1 to 5526)
REFERENCE      Entcheva, P., Liebl, W. and Streit, W.R.
AUTHORS      Direct Submision
TITLE      Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
          Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
FEATURES
source
1. 5526
/organism="uncultured bacterium pCosHE2"
/db_xref="taxon:143797"
/clone="pCosHE2"
/note="unknown organism, cosmid clone derived from
          environmental consortium"
complement(52..528)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="hypotheical 17.1 kDa protein in modc-bioc
          intergenic region"
gene
CDS
/protein_id="AAG60577.1"
/db_xref="GI:12620125"
/translation="MKLISNDLRDDCKLPBHRHVENGMGVDGDNISPHLAWDDVPAGTK
          SVVAVYDADATGSGMWHVYVNLPAFTRVLPFGSGGLVAMPDGYLOFTPDGKAG
          YGCAAPRPGFHRITYFYHADLIEIDVDEGASGAMGVFNHPSLASASTTAMFS"
complement(587..1876)
/gene="bioA"
/note="7.8-diaminopelargonic acid
          synthetase-aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase BioA"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDDLAEPDORHITWHPYTSMTSPPLPVYVVASGCELLISDGR
          LVDGSSWMAIHGYNHNPOLMAAKMSQIDAMSHYFPGTITPARAIELCRKIVYTPOP
          LKCVLADSGSYAVAVAKMAILOWAKGEARQFRTFRNGYHDTGTGASVCDPDS
          MSLMKGYLPENLPAPOSRMDDEBERDQVGFARLMAARHIEIAVILPEIVGAG
          GMRMYHPMLKRIKICDREGILITADEIATFGRTGKLFACHEAIEIAPDILGKAL
          TCGWTLSATLTREVAETISNGAGCGMHGPTFMGNPLACAAANASIALIESDMOO
          QVADIEVOLREOLAPARDAEMVADVRLVIGAVETTHPVNMAALQKFEVQGVWIRP
          FGLIYIMPYIILPQOLRLTAANNVAVDETFEQ"
1963..3003
/gene="bioB"
1963..3003
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRPMTLSQVTELEFKPLDLLFEAQVHROHPDPROYVST
          LISIKTGACPEDECKYCDQSSRYKTGLAEERIMEVEQVLESARAKAKAGTRFCGAM
          KNPNRDMPLYLEQWOGVKGDLGACMTGLTSSQORLANAGLDYNNILDPPEF
          YGNITTRTYOERLDITLEKVRDAGIKVCSGIVLIGTVDRGLLIOLLNLPSPS
          VPIMVLYVKGTFPLANDNDPAARFETITIANARIMPTSVYRSAGSEQNEQIAMC
          PMAGANSFTFYCKLITTPNPEKDIDLFRLGLNPOOTVAVLGDNEQDQRLDQALMT
          PTDYVYMAAL"
3000..4154
/gene="bioF"
3000..4154
/gene="bioF"
/note="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase BioF"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSWOKRTINALDRAADALRRRYPAAGAGRWLVADDOYVNF
          SSNDYLSHHPQIIRAMOGAGDEFGSGSGSHVSGSYVYHDALEELAEWIGISRA
          LLFTSGFRANAVIAMAAMAKEDRIAURLSHASLSEASLSPQLRFAHNDVTHAR
          LLASPCPGQOAVITEGVSMDSAPLAEIQVYQDNGWMLMDADAGTGVIEQDGRG
          SCWLQKVPDLVVTFGKGVSGAVALCSTVADYLLQFARHLIYSTSMPPAQOAL
          RASLAVIRDEGDARREKLIVLAFHFAVQODLPFTLADSCSAIOLIVDNGSRALOL
          AEKLRQOCQWTAIRPTVPAGTARLTLTAHEMDQDIDRLLEVLHGNC"
4141..4896
/gene="bioc"
4141..4896
/gene="bioc"
/note="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein bioc"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVNRQALAAAFGRRAALYEQHADLQORSADVLLAMLPERRY
          HVLDAGCGPMWTRHMRERHAQVYATADLSPMITVOAROKDAADHYLAGDIESLPLTA
          TFDLAMSNLAVOWCGNISTALRELYRVRSKGVAFATTVQSGSLPELHOMQAVDPR
          HANFELPDELEOSLNGVYOHNIOPITLTFDDALSAMRSIKGATGTHLHGSDPRIL
          TRSRLORLQALAMPQOQGRPLYTHLPLVGIARE"
4883..5526

```



CDS

/gene="blod"  
4883..>5526  
/gene="blod"  
/note="DDB synthetase"  
/codon\_start=1  
/transl\_table=11  
/product="dehydrobiotin synthetase Blod"  
/protein\_id="AAG60582.1"  
/db\_xref="GI:12620130"

BASE COUNT 1274 a 1507 c 1567 g 1178 t

ORIGIN

Query Match 60.0%; Score 15; DB 2; Length 5526;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgaatcctgcagct 15  
|||||

Db 2067 CGATCCTGCTCAGGT 2081

RESULT 43  
AF237379 8326 bp DNA VRT 13-NOV-2000  
LOCUS Xenopus laevis neurofilament protein (NF-M(2)) gene, complete cds.  
ACCESSION AF237379  
VERSION AF237379.1 GI:11139431  
KEYWORDS  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Xenopus.  
1 (bases 1 to 8326)  
Roosa, J.R., Gervasi, C. and Szaro, B.G.  
Structure, biological activity of a middle molecular mass neurofilament gene  
of Xenopus laevis  
Brain Res. Mol. Brain Res. 82 (1-2), 35-51 (2000)  
JOURNAL 11042356  
PUBMED 2 (bases 1 to 8326)  
AUTHORS Roosa, J.R. and Szaro, B.G.  
TITLE Direct Submission  
JOURNAL Submitted (18-FEB-2000) Biological Sciences, University at Albany,  
State University of New York, 1400 Washington Avenue, Albany, NY  
12222, USA

FEATURES

source Location/Qualifiers  
1..8326  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
TATA\_signal 1498..1503  
5'UTR 1524..1563  
exon 1524..2598  
/number=1  
join(1524..2598,3533..3657,6507..>8326)  
/product="neurofilament protein"  
1564..>8326  
/gene="NF-M(2)"  
join(1564..2598,3533..3657,6507..8088)  
/gene="NF-M(2)"  
/codon\_start=1  
/product="neurofilament protein"  
/protein\_id="AAG31697.1"  
/db\_xref="GI:11139432"  
/translation="MYSLEFGNPSYRMAETPTYSRPTSSPSSGFRSQRSSRP  
TTTSSKRTNLAAPRVSSNDSLDSSSAFNGDLKQVRSSEKQDLGDRVGLYE  
KVHLEQNKTEIAETALRKQSGISQLEVEVEQERKEELRVNLELGHDKAIVLDT  
EHLDDIQRLEERLEEARIRDTEAAIRAMKDMEDTSLVKKVELDKVSLQEEVAF

INTRON

/gene="NF-M(2)"  
/number=1  
3533..3657  
/gene="NF-M(2)"  
/number=2  
3659..6507  
/gene="NF-M(2)"  
/number=2  
6507..>8326  
/gene="NF-M(2)"  
/number=3  
8089..>8326  
/gene="NF-M(2)"

BASE COUNT 2631 a 1560 c 2161 g 1974 t

ORIGIN

Query Match 60.0%; Score 15; DB 8; Length 8326;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caggtgcagtcagc 25  
|||||

Db 3060 CAGGTGCAGTCAGC 3074

RESULT 44  
AE004708 12542 bp DNA BCT 30-AUG-2000  
LOCUS Pseudomonas aeruginosa PA01, section 269 of 529 of the complete  
genome.  
DEFINITION AE004708 AE004091  
ACCESSION AE004708.1 GI:9948890  
VERSION  
KEYWORDS  
SOURCE Pseudomonas aeruginosa.  
ORGANISM Pseudomonas aeruginosa.  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
Pseudomonas.  
1 (bases 1 to 12542)  
Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P.,  
Hickey, M.J., Brinkman, F.S., Huftagle, W.O., Kowalik, D.J., Lagrou, M.,  
Garber, R.L., Goltzy, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y.,  
Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,  
Smith, K., Spencer, D., Wong, G.K., Wu, Z. and Paulsen, I.T.  
Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen  
Nature 406 (6799), 959-964 (2000)  
JOURNAL 20437337  
MEDLINE 2 (bases 1 to 12542)  
AUTHORS Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,  
Hickey, M.J., Brinkman, F.S.L., Huftagle, W.O., Kowalik, D.J., Lagrou, M.,  
Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E.,  
Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,  
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,  
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H.,  
Hancock, R.E.W., Lofly, S. and Olson, M.V.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA

| FEATURES |          | Location/Qualifiers  |
|----------|----------|--|
| source   | 1..12542 | /organism="Pseudomonas aeruginosa"                           |
|          |          | /strain="PA01"   |
| gene     |          | /db_xref="taxon:287"   |
|          |          | /complement(90..707)   |
| CDS      |          | /gene="PA2807"   |
|          |          | /complement(90..707)   |
|          |          | /gene="PA2807"   |
|          |          | /codon_start=1   |
|          |          | /transl_table=11   |
|          |          | /product="hypothetical protein"                              |
|          |          | /protein_id="AAG06195.1"                                     |
|          |          | /db_xref="GI:9948891"  |
|          |          | /translation="MLPTASROAQRHIGDIMPMEIPRMFPRLLPASLITVGLVGA      |
|          |          | SAOASPAHGOAFGKPAOASRSIEVVLGDMYFKPRAIEVAGETVRFVLEKEKLL        |
|          |          | HEENLGDAMAHAEHOKEMLEMOOSGMLPTGMAHDSHOSMGAMDMDHGRMKKDDP       |
|          |          | NSVLYPRGKSALETWTFKATRLERACNIPGHYQAMGQLTVP"                   |
| gene     |          | /complement(879..1070)                                       |
| CDS      |          | /gene="PA2808"   |
|          |          | /complement(879..1070)                                       |
|          |          | /gene="PA2808"   |
|          |          | /codon_start=1   |
|          |          | /transl_table=11   |
|          |          | /product="hypothetical protein"                              |
|          |          | /protein_id="AAG06196.1"                                     |
|          |          | /db_xref="GI:9948892"  |
|          |          | /translation="MRTFTASTLAPLLVLPVPMADSGDIVFERMOQAARAA          |
|          |          | HOEKLAKOAEATRKDKENC"   |
| gene     |          | 1196..1876   |
|          |          | /gene="PA2809"   |
| CDS      |          | 1196..1876   |
|          |          | /gene="PA2809"   |
|          |          | /codon_start=1   |
|          |          | /transl_table=11   |
|          |          | /product="probable two-component response regulator"         |
|          |          | /protein_id="AAG06197.1"                                     |
|          |          | /db_xref="GI:9948893"  |
|          |          | /translation="MKLLIVEDEPRIGQVLRQGLAEAGRAVLDSDNGNEGEOALGCD    |
|          |          | YDLILIDVMLPGRDGMQILRSYVDAGMTVPVLTARQDVEDRVRKLEGGADYLVKP      |
|          |          | FAFVELARVRLTLRRGSOQLQETTLQIADLELDILRRVQRQGRIDLTIAEFALLE      |
|          |          | LILRRSGEVLPRSLIASQVDMNFDSDTVIEVALIRLRAKVDDDDYPOQLITVGMG      |
|          |          | YLEERDE"   |
| gene     |          | 1873..3204   |
|          |          | /gene="PA2810"   |
| CDS      |          | 1873..3204   |
|          |          | /gene="PA2810"   |
|          |          | /codon_start=1   |
|          |          | /transl_table=11   |
|          |          | /product="probable two-component sensor"                     |
|          |          | /protein_id="AAG06198.1"                                     |
|          |          | /db_xref="GI:9948894"  |
|          |          | /translation="MSAGFGRSMSCVGRSLFPACTAAVSLIAGLIISRDAIDHEVE     |
|          |          | LDMAMSAKLAVRDELKGLGSEQMRREAEELRLARHPDLGLNLGPDNLWFER          |
|          |          | LDPAPHGUPANRELGAFLPEPNDASPRLVIIIDISHQFLQRMROLIWMISJAL        |
|          |          | ATALLGAMATGASLAPLRMRREVAARANSLLTRLDASRMEBELGLGELNAMIAR       |
|          |          | LEAPQRLSAFSAHDELRTPLTSLTQOVYLSRSLDEYREALHGNLELELT            |
|          |          | AMVMDMLIAKADHGLIAPSRQLDGAEVDSLEFYOPLAEDRDIRLIRBSLSIPG        |
|          |          | DRGMRLRVIANLIDNALRTTADGGERIRIRIGDRRLSVEOGPAIIPERLPRLEDRYR    |
|          |          | ADPARREGOEHEAGLAIQCRSTVQAHGSGIRCESADGWTREVIDARPRR"           |
| gene     |          | /complement(3254..4033)                                      |
| CDS      |          | /gene="PA2811"   |
|          |          | /complement(3254..4033)                                      |
|          |          | /gene="PA2811"   |
|          |          | /codon_start=1   |
|          |          | /transl_table=11   |
|          |          | /product="probable permease of ABC-2 transporter"            |
|          |          | /protein_id="AAG06199.1"                                     |
|          |          | /db_xref="GI:9948895"  |
|          |          | /translation="MSSELSANLVALNTIYRVRRTLRIMPTLLPPIATMYLYVI       |
|          |          | FGNLIGROIGDMGSGTYMEYIVPGLIMSVITAYGAVSSPFGSGEORSVEELIASP      |
|          |          | VSPHTILAGTIGVLRGLAVGVYVIMLSLFTKLOYHHGITYLYVYVLTATITFSLGG     |
|          |          | FINAIVFRNFDIISILPTFVLTPLTYLGGVFYSINLLPFWQTVSLANLILHVNFR      |
|          |          |  |
| gene     |          | YGLIGSDIRICVAGFMILATAVLYGCVRLLYSGRGMRQ"                      |
| CDS      |          | /complement(4030..4962)                                      |
|          |          | /gene="PA2812"   |
|          |          | /complement(4030..4962)                                      |
|          |          | /gene="PA2812"   |
|          |          | /codon_start=1   |
|          |          | /transl_table=11   |
|          |          | /product="probable ATP-binding component of ABC transporter" |
|          |          | /protein_id="AAG06200.1"                                     |
|          |          | /db_xref="GI:9948896"  |
|          |          | /translation="MSGALSIRQLTKTYGNGFOALGIDDDVAEGDFEALLGPNAGK     |
|          |          | STTGILSTLVNKSQSVSGYGDLDKDPAGIKRCIGVQGEENFMQFEKPIVITO         |
|          |          | AGYIGITPAKTAKERARYITQGLMDKRNASRLSGMKRRRLIMAIRAIVHORILL       |
|          |          | DEPRAVDIELRRESMSFTELVQNEGISTILLTHYLEAQDLQRLNAILIDHETVQNT     |
|          |          | SMRILMTLHSEFLLDLKVVQALPPLDGYPRRLVDHDTLVEVERKSQGINDLFAOL      |
|          |          | GAQGIETLSRNKTNRLLEELFVSILVERLTRIAR"                          |
| gene     |          | /complement(5038..5658)                                      |
| CDS      |          | /gene="PA2813"   |
|          |          | /complement(5038..5658)                                      |
|          |          | /gene="PA2813"   |
|          |          | /codon_start=1   |
|          |          | /transl_table=11   |
|          |          | /product="probable glutathione S-transferase"                |
|          |          | /protein_id="AAG06201.1"                                     |
|          |          | /db_xref="GI:9948897"  |
|          |          | /translation="MLTIKGRKSSNRYKALMCAEEAGLEFYRTVDAGARGVLDPEAP    |
|          |          | RAMPNRVRVIEDGDFVLDSNAIVYTLARAPGLYFQDYVRADADKMDWTTS           |
|          |          | TLAPFRDLFWGLTRTPPEQDEALAKALQTCGELLRVDDTLAQGPWLISGRFGWG       |
|          |          | DIPLGCFIVAFMEPIERPPLHLEAWYLRERPARTAVMELT"                    |
| gene     |          | /complement(5773..6444)                                      |
| CDS      |          | /gene="PA2814"   |
|          |          | /complement(5773..6444)                                      |
|          |          | /gene="PA2814"   |
|          |          | /codon_start=1   |
|          |          | /transl_table=11   |
|          |          | /product="hypothetical protein"                              |
|          |          | /protein_id="AAG06202.1"                                     |
|          |          | /db_xref="GI:9948898"  |
|          |          | /translation="MOSYLSGRVDSHDHPVYVEFAKRSNKAQPRDOVALYAVR        |
|          |          | DGVRNIPVFSRDPQTLKASALIDQGSSTYCPKAILAACARRCRIPARIGLDVRH       |
|          |          | LATPRLLLEALSEVFAMHGVTYELTEBRWKATPAFRALCRAPDAVLEPFDVAOSV      |
|          |          | FHPNRQGERMYEYLADHGFADLPEELFESHQOHYPLHESGRPLALDDGFOAEAGQ      |
|          |          | DEGRR"   |
| gene     |          | 6597..9044   |
| CDS      |          | /gene="PA2815"   |
|          |          | 6597..9044   |
|          |          | /gene="PA2815"   |
|          |          | /codon_start=1   |
|          |          | /transl_table=11   |
|          |          | /product="probable acyl-CoA dehydrogenase"                   |
|          |          | /protein_id="AAG06203.1"                                     |
|          |          | /db_xref="GI:9948899"  |
|          |          | /translation="MLLMVLYLVCAVLAHRTTPAPALGISAATLIIMGVESHAP       |
|          |          | GWLLLVFEMLVLAVALPLADLRRVLSGLFWFQKVLPPMSQTERETREAGTVM         |
|          |          | DGEELFSPDMQKLLDPRKALTEEDQAFVGPPELCAVMSMDIGORDLDPEAAV         |
|          |          | FTKQHGFEGLIIPKEYGKGSARAHQVYMKLATRSGDLASTVPMVNSGSPPELLLT      |
|          |          | YGTDEORGRVYLSLAKGDDIPCFALTYAGSDGMDYGVYVCKGMEBEREVLGLTL       |
|          |          | TMKRYITIGVATVLLIAGFCHPDPLIDGDEEDGITLAILPDPFGVEIGRRHYL        |
|          |          | GAAPNGNSCKGVDVPLEYITIGGQEMIGKMMMLNMCISVGRSISLPAVTSAGKRS      |
|          |          | SVSGRTAQRVEQRFNVPAAEFQIOWALARIQGNAMLMDSARITTAADVLSKPSYL      |
|          |          | SAILKYLHTEGRRECIAMADIHGGKGIKGPNNYLDKSMOGATITTVYEGANILSN      |
|          |          | LMIFGOGAIRCHPVYVIAKEMELADREDDQAREPDALLVYHIGFAVNAASSFLTSS     |
|          |          | CGHLGNAPGDRISRPYFRALNRLAASFALADFSMLLGGELKRELSARLGDVLSY       |
|          |          | LYLGSAAKLRHYHDGNPDYLPPLIRMAEMEMLGKAEMALIEDLLNPFSPFGCALKYL    |
|          |          | VLPRGRRRHGGDELDIAETILGRPDPPDPAOLAILGAFLPKDDPODPVGLAAAFDA     |
|          |          | VRESALEKTIHKAKIKBERVOPQAGNPIDAGAAAGVLSAEOALALHQAERARKKVID    |
|          |          | VDEKKEELQADGKIR"   |
| gene     |          | 9173..9553   |
| CDS      |          | /gene="PA2816"   |
|          |          | 9173..9553   |
|          |          | /gene="PA2816"   |
|          |          | /codon_start=1   |

Search completed: October 9, 2001, 15:48:27  
 Job time: 14893 sec

```

/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06204.1"
/db_xref="GI:9948900"
/translation="MLKFRATGLASFLIPPAQAAPPAFYGVPHQAQAVORAGEQRO
KQLQEDQRFDEQRLQLODQLOQROONLQROQROMODNLRQQLDQQRWRLQED
QRLDSERRQLENRRROSQSPAIR"
9612..10022
/gene="PA2817"
9612..10022
/gene="PA2817"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06205.1"
/db_xref="GI:9948901"
/translation="MANSTLEHHLALLQHLRNILGALGEAEQVPEENHGLFLERDEL
MDLPRDPGAGYLGDLLISOVHRYPTALHLPRLDMFPGDCLHFMPEDEIDLFQ
KLDERRYEAEERGETFDMNRERQLALPDPSPRH"
complement(10089..11666)
/gene="PA2818"
complement(10089..11666)

```

Query Match 60.0%; Score 15; DB 1; Length 12542;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgaatcctgcaggt 15  
 |||  
 Db 4038 CGATCCTGTCAGGT 4052

## RESULT 45

AC014847 AC014847 20780 bp DNA HTG 16-NOV-1999  
 LOCUS Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
 DEFINITION pieces.  
 AC014847  
 AC014847.1 GI:6436488  
 VERSION HTG: HTGS\_PHASE2.  
 KEYWORDS fruit fly.  
 SOURCE  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 20780)  
 Authors Adams, M. and Venter, J.C.  
 Title Direct Submission  
 Journal Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA

COMMENT  
 This sequence was identified as CDM:10211717 by the submitter.  
 For further information on this sequence e-mail to fly@celera.com.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

FEATURES  
 source  
 1..20780  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"

BASE COUNT 5541 a 4541 c 4915 g 5783 t  
 ORIGIN

Query Match 60.0%; Score 15; DB 63; Length 20780;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 tgcgcaggtgcaggt 21  
 |||  
 Db 17291 TCGTCAGGTGTCAGGT 17305

Wed Oct 10 07:46:11 2001

us-09-396-196f-9.oli.rge

Page 38

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:53:10 ; Search time 469.84 Seconds  
(without alignments)  
33.410 Million cell updates/sec

Title: US-09-396-196F-9

Perfect score: 25

Sequence: 1 cgatcctcgtcaggtgcagtcagc 25

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 32919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq.0601.\*

- 1: /SIDSI/gcgcdata/geneseq/geneseqn/NA1980.DAT.\*
- 2: /SIDSI/gcgcdata/geneseq/geneseqn/NA1981.DAT.\*
- 3: /SIDSI/gcgcdata/geneseq/geneseqn/NA1982.DAT.\*
- 4: /SIDSI/gcgcdata/geneseq/geneseqn/NA1983.DAT.\*
- 5: /SIDSI/gcgcdata/geneseq/geneseqn/NA1984.DAT.\*
- 6: /SIDSI/gcgcdata/geneseq/geneseqn/NA1985.DAT.\*
- 7: /SIDSI/gcgcdata/geneseq/geneseqn/NA1986.DAT.\*
- 8: /SIDSI/gcgcdata/geneseq/geneseqn/NA1987.DAT.\*
- 9: /SIDSI/gcgcdata/geneseq/geneseqn/NA1988.DAT.\*
- 10: /SIDSI/gcgcdata/geneseq/geneseqn/NA1989.DAT.\*
- 11: /SIDSI/gcgcdata/geneseq/geneseqn/NA1990.DAT.\*
- 12: /SIDSI/gcgcdata/geneseq/geneseqn/NA1991.DAT.\*
- 13: /SIDSI/gcgcdata/geneseq/geneseqn/NA1992.DAT.\*
- 14: /SIDSI/gcgcdata/geneseq/geneseqn/NA1993.DAT.\*
- 15: /SIDSI/gcgcdata/geneseq/geneseqn/NA1994.DAT.\*
- 16: /SIDSI/gcgcdata/geneseq/geneseqn/NA1995.DAT.\*
- 17: /SIDSI/gcgcdata/geneseq/geneseqn/NA1996.DAT.\*
- 18: /SIDSI/gcgcdata/geneseq/geneseqn/NA1997.DAT.\*
- 19: /SIDSI/gcgcdata/geneseq/geneseqn/NA1998.DAT.\*
- 20: /SIDSI/gcgcdata/geneseq/geneseqn/NA1999.DAT.\*
- 21: /SIDSI/gcgcdata/geneseq/geneseqn/NA2000.DAT.\*
- 22: /SIDSI/gcgcdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 25    | 100.0       | 1041   | 20    | AA01303     |
| 2          | 25    | 100.0       | 1084   | 10    | AA01329     |
| 3          | 25    | 100.0       | 1121   | 7     | AA060496    |
| 4          | 25    | 100.0       | 5872   | 15    | AA062386    |
| 5          | 15    | 60.0        | 34     | 16    | AA075832    |
| 6          | 15    | 60.0        | 353    | 21    | AA056641    |
| 7          | 15    | 60.0        | 586    | 21    | AA075563    |
| 8          | 15    | 60.0        | 4674   | 20    | AA090421    |
| 9          | 15    | 60.0        | 327    | 21    | AA000048    |
| 10         | 14    | 56.0        | 374    | 22    | AA064981    |
| 11         | 14    | 56.0        | 423    | 14    | AA045597    |

#### ALIGNMENTS

| Result ID | Score | Query Match | Length | DB ID | Description |
|-----------|-------|-------------|--------|-------|-------------|
| 1         | 25    | 100.0       | 1041   | 20    | AA01303     |
| 2         | 25    | 100.0       | 1084   | 10    | AA01329     |
| 3         | 25    | 100.0       | 1121   | 7     | AA060496    |
| 4         | 25    | 100.0       | 5872   | 15    | AA062386    |
| 5         | 15    | 60.0        | 34     | 16    | AA075832    |
| 6         | 15    | 60.0        | 353    | 21    | AA056641    |
| 7         | 15    | 60.0        | 586    | 21    | AA075563    |
| 8         | 15    | 60.0        | 4674   | 20    | AA090421    |
| 9         | 15    | 60.0        | 327    | 21    | AA000048    |
| 10        | 14    | 56.0        | 374    | 22    | AA064981    |
| 11        | 14    | 56.0        | 423    | 14    | AA045597    |

RESULT 1  
AA01303 standard: DNA; 1041 BP.

AA01303:

12-APR-1999 (first entry)

E. coli biotin synthetase (Biot) coding sequence.

DAP aminotransferase; diaminopelargonic acid; transgenic plant;

biotin synthetase; biotin production; vitamin H; Biot; ss.

Escherichia coli.

US5869719-A.

09-FEB-1999.

30-APR-1997; 97US-0846338.

30-APR-1997; 97US-0846338.

08-MAR-1995; 95US-0401068.

(NOVS) NOVARTIS FINANCE CORP.

Patton DA;

WPI: 1999-152902/13.

P-PSDB; AA073906.

Transgenic plants with high biotin levels - transformed with DNA

encoding diaminopelargonic acid amino-transferase or biotin synthase

XX Example 2: Column 37-40; 34pp: English.

CC This sequence encodes the E. coli biotin synthetase (Biotin). The gene can  
CC be used in the transgenic plant of the invention. The transgenic plant,  
CC plant cell or plant tissue is transformed with a chimeric gene encoding  
CC diaminopelargonate acid (DAP) aminotransferase or biotin synthase and  
CC produces more biotin than a non-transgenic plant, cell or tissue. The  
CC plant is used as an improved dietary source of biotin (vitamin H) for  
CC humans or animals.

XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 0.00015;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtcaggtgcagtcagc 25  
|||||  
Db 105 cgatcctcgtcaggtgcagtcagc 129

RESULT 2

AA091329 standard; DNA; 1084 BP.

AA091329;

15-FEB-1990 (first entry)

E.coli Bio B gene.

E.coli Bio B gene; biotin.

Escherichia coli.

Key Location/Qualifiers

CDS 24..1064

GB2216530-A.

11-OCT-1989.

17-MAR-1989; 89GB-0006210.

22-MAR-1988; 88GB-0006804.

17-MAR-1989; 89GB-0006210.

(UKAG-) UK MIN. AGRIC. FISH.

Pearson BM, McKee RA;

WPI; 1989-295085/41. P-PSDB P91392

Plasmid contg. gene(s) for expression of biotin synthetase enzymes  
- derived from E.coli and capable of replication and expression in other  
microorganisms, esp. yeast.

Table 3; page 33-4; 52pp: English.

The gene can be used in a plasmid for expression of enzymes of the biotin  
synthetic pathway. Pref. control sequences for expression in S.cerevisiae  
are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for  
lactobacillus. Insertion of bio B improves biotin yields in  
microorganisms which export biotin, or enables growth in media contg.  
little or no biotin of organisms unable to synthesise biotin for their  
own use.

Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;  
Best Local Similarity 100.0%; Pred. No. 0.00014;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtcaggtgcagtcagc 25  
|||||  
Db 128 cgatcctcgtcaggtgcagtcagc 152

RESULT 3

AA060496 standard; DNA; 1121 BP.

AA060496;

17-OCT-1991 (first entry)

Sequence encoding biotin synthesising enzyme.

Biotin synthetic enzyme; E.coli; desthiobiotin; ds.

Key Location/Qualifiers

CDS 42..1082

JP61149091-A.

07-JUL-1986.

24-DEC-1984; 84JP-0272605.

24-DEC-1984; 84JP-0272605.

(NIPS ) NIPPON SODA KK.

WPI; 1986-216622/33.

P-PSDB; AAP60536.

Double stranded DNA encoding biotin synthesising enzyme -  
comprises transformed mutant E.coli strain contg. cyclic doubled  
stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

Disclosure; Page 534; 23pp: Japanese.

The sequence may be expressed by a transformed E.coli host, cultured  
in a medium containing desthiobiotin.

Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;  
Best Local Similarity 100.0%; Pred. No. 0.00014;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtcaggtgcagtcagc 25  
|||||  
Db 146 cgatcctcgtcaggtgcagtcagc 170

RESULT 4

AA062386 standard; DNA; 5872 BP.

AA062386;

16-NOV-1994 (first entry)

Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

Biotin; expression; enterobacteria; vitamin H; synthesis;  
plasmid; pB030A-15/9; bioB; bioC; bioD; bioA;

promoter; pta; biotin synthase; KAPA synthase;  
8-amino-7-oxononanoate synthase; pimeoyl-CoA; DTB synthase;

```

KM dethiobiotin synthase; DAPA synthase;
KM S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;
XX seborrhoea; dermatitis; ds.
OS Escherichia coli DSM498.

XX Key Location/Qualifiers
XX promoter 1..96
FT /tag= a
FT /function= "promoter plac"
FT /evidence= EXPERIMENTAL
FT 23..28
FT /tag= b
FT /standard_name= "promoter plac"
FT 45..50
FT /tag= c
FT /evidence= EXPERIMENTAL
FT /standard_name= "promoter plac"
FT 105..109
FT /tag= d
FT /evidence= EXPERIMENTAL
FT /standard_name= "bioB RBS no. 9"
FT 117..1157
FT /tag= e
FT /product= "biotin synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioB"
FT /number= 1
FT 1141..1146
FT /tag= f
FT /standard_name= "bioF RBS"
FT 1154..2311
FT /tag= g
FT /EC_number= 2.3.1.47
FT /product= "KAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioF"
FT /number= 2
FT /standard_name= "8-amino-7-oxononanoate synthase"
FT 2284..2288
FT /tag= h
FT /standard_name= "bioC RBS"
FT 2295..3050
FT /tag= i
FT /function= "involved in pimeloyl-CoA synthesis"
FT /product= "protein"
FT /gene= "bioC"
FT /number= 3
FT 3030..3033
FT /tag= j
FT /standard_name= "bioD RBS"
FT 3043..3753
FT /tag= k
FT /EC_number= 6.3.3.3
FT /product= "PTB synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioD15"
FT /number= 4
FT /standard_name= "dethiobiotin synthase"
FT 3712..3750
FT /tag= l
FT /note= "bioD15 substitution"
FT 3742..3746
FT /tag= m
FT /standard_name= "bioA RBS"
FT 3750..5039
FT /tag= n
FT /EC_number= 2.6.1.62
FT /product= "DAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioA"
FT /number= 5
FT /standard_name= "S-adenosyl-L-methionine; 8-amino-

```

```

FT RBS 5088..5093 7-oxononanoate aminotransferase"
FT /tag= o
FT /standard_name= "ORF1 RBS"
FT 5098..5574
FT /tag= p
FT /function= "unknown, involved in biotin synthesis"
FT /product= "protein"
FT /evidence= EXPERIMENTAL
FT /gene= "ORF1"
FT /number= 6
FT 5583..5644
FT /tag= q
FT /standard_name= "rho-independent transcriptional
FT terminator"
FT stem_loop 5583..5605
FT /tag= r
FT MO9408023-A.
FT 14-APR-1994.
FT 01-OCT-1993; 93WO-EP02688.
FT 02-OCT-1992; 92CH-0003124.
FT 15-JUL-1993; 93CH-0002134.
FT (LONZ ) LONZA AG.
FT Birch O, Brass J, Fuhrmann M, Shaw N;
FT WPI: 1994-135587/16.
FT P-FSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT Biotechnological biotin prodn. using enterobacterial biotin-gene
FT - providing vitamin H in high yield
FT Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
FT The sequence is derived from plasmid pB030A-15/9 contg. the
FT bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FT of biotin, arranged in a transcription unit. Microorganisms
FT contg. these DNA fragments or plasmids may be used in the prodn.
FT of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FT loss of appetite and tiredness.
FT Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
FT
FT Query Match 100.0%; Score 25; DB 15; Length 5872;
FT Best Local Similarity 100.0%; Pred. No. 0.00013;
FT Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT QY 1 cgatcctgcagtcagtcagtcagc 25
FT |||||
FT Db 221 cgatcctgcagtcagtcagtcagc 245
FT
FT RESULT 5
FT ID AA075832
FT ID AA075832 standard; DNA; 34 BP.
FT AC AA075832;
FT DT 18-AUG-1995 (first entry)
FT DE Sense primer to amplify Non-A Non-B hepatitis virus for analysis.
FT XX Non-A Non-B hepatitis virus; structural region; cDNA to genomic RNA;
FT KM detection; reagent; anti-Non-A Non-B hepatitis virus antibody;
FT KM vaccine; antigen; epitope; diagnosis; ss.
FT XX Synthetic.
FT OS

```

```

XX  EP628572-A.
PN
XX  14-DEC-1994.
PD
XX
XX  27-MAY-1994: 94EP-0108256.
PF
XX
XX  28-MAY-1993: 93JP-0126709.
PR
XX  02-MAR-1994: 94JP-0032201.
PA
XX  (ARIM/) ARIMA T.
XX  (EISA ) EISAI CO LTD.
PI
PI  Aoyama M, Arima T, Hosoda T, Iwasaki Y, Obara T,
PI  Sawada T, Tohmatsu J;
DR  WPI; 1995-015655/03.
XX
XX  New non-A non-B hepatitis virus sub-type - used to develop prods.
PT  for detection, diagnosis, prevention and treatment of non-A non-B
PT  hepatitis.
XX
XX  Example 2; Page 49; 59pp; English.
XX
XX  This primer is based on nucleotides 778-800 of AAQ75818 (a part of
XX  the Non-A Non-B hepatitis virus genome encoding the non-structural
XX  protein). It is used in conjunction with AAQ75833 to amplify nucleotides
XX  800-1326 of AAQ75818. The nucleotide sequences (see also AAQ75817-19)
XX  were isolated from the plasma of donors in Japan with high s-GFP levels,
XX  and were found to be different from previously reported NANB hepatitis
XX  viruses. The DNA can be used as a reagent for detecting the NANB
XX  hepatitis viral gene. The polypeptides can be used as reagents for
XX  detecting anti-NANB hepatitis antibodies or as a NANB hepatitis viral
XX  vaccine.
SQ  Sequence 34 BP; 6 A; 9 C; 12 G; 7 T; 0 other;

Query Match      60.0%; Score 15; DB 16; Length 34;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  2 gatccctgcagtg 16
    |||||||||||||
DB  8 gatccctgcagtg 22

RESULT 6
AAC56641/c
ID  AAC56641 standard; DNA; 353 BP.
XX
XX  AAC56641;
AC
XX
XX  25-JAN-2001 (first entry)
DT
XX
XX  Eucalyptus grandis transcription factor DNA sequence #512.
DE
XX
XX  Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
XX  poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
XX  basic helix-loop-helix zipper; homeotic; homeobox; MADS;
XX  homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
XX  type 2 Cys2His2; CCAAT box element; MYB; ss.
OS  Eucalyptus grandis.
XX
XX  WO200053724-A2.
PN
XX
XX  14-SEP-2000.
PD
XX
XX  09-MAR-2000; 2000WO-US06112.
PF
XX
XX  11-MAR-1999; 99US-0266513.
PR
XX  18-AUG-1999; 99US-0149485.

```

```

XX  (GENE-) GENESIS RES & DEV CORP LTD.
PA  (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX  Wood M, McGrath A, Shenk MA, Glenn M;
PI  WPI; 2000-579369/54.
DR
XX
XX  New isolated polynucleotide encoding a plant transcription factor for
PT  producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT  having modified gene expression or modified activity of a polypeptide
XX
XX  Claim 1; Page 482; 747pp; English.
XX
XX  The present invention relates to novel plant transcription factors from
XX  Eucalyptus grandis or Pinus radiata. The present sequence is the coding
XX  sequence for one such transcription factor. The transcription factor may
XX  be used to produce a plant having modified gene expression such as a
XX  woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
XX  mahogany species or to modify the activity of a polypeptide in a plant.
XX  The transcription factors of the present invention are members from the
XX  following families of regulatory proteins: bZIP, bZIP family of G-box
XX  binding factors, basic helix-loop-helix zipper, LIM domain, AP2
XX  homeotic/homeodomain/homeobox/MADS, homeobox zipper, LIM domain, AP2
XX  and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
XX  and MYB.
SQ  Sequence 353 BP; 117 A; 59 C; 98 G; 78 T; 1 other;

Query Match      60.0%; Score 15; DB 21; Length 353;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  7 tccatcagtgaggt 21
    |||||||||||||
DB  326 TCGTCAGGTGCAAGT 312

RESULT 7
AAC75563/c
ID  AAC75563 standard; cDNA; 586 BP.
XX
XX  AAC75563;
AC
XX
XX  08-FEB-2001 (first entry)
DT
XX
XX  Human ORF118 polynucleotide sequence SEQ ID NO:2235.
DE
XX
XX  Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
XX  vulnery; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
XX  anticonvulsant; osteoprotic; antiarthritic; immunosuppressant; cardiac;
XX  immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX  hypotensive; dermatological; immunosuppressive; antineoplastic;
XX  antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
XX  antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
XX  neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX  cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX  cholesterol ester storage; systemic lupus erythematosus; infection;
XX  severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX  allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX  bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX  thrombosis; contraceptive; ss.
OS  Homo sapiens.
XX
XX  WO200058473-A2.
PN
XX
XX  05-OCT-2000.
PD
XX
XX  31-MAR-2000; 2000WO-US08621.
PF
XX
XX

```



PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CUBA-) CURAGEN CORP.  
 XX  
 PI Shinkets RA, leach M;  
 XX  
 DR WPI: 2000-602362/57.  
 DR P-PSDB: AAB41354.  
 XX  
 PS Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 5; Page 1649; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antiproliferative; antiparasitism; nootropic; neuroprotective;  
 CC osteoprotective; anticonvulsant; antirheumatic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasodilator;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antineoplastic; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antihypertensive; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC erythematous, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 586 BP; 122 A; 165 C; 156 G; 123 T; 0 other;

Query Match 60.0%; Score 15; DB 21; Length 586;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atcctcgtcaggtgc 17  
 |||||  
 DB 304 ATCCTGTCAGGTGC 290

RESULT 8  
 AAX90421/c  
 ID AAX90421 standard; cDNA; 4674 BP.  
 XX  
 AC AAX90421;  
 XX  
 DT 29-SEP-1999 (first entry)  
 XX  
 DE Human ataxin-2 like gene.  
 XX  
 KM Human: ataxin-2 like gene; ataxin-2 like protein; ataxia; deafness;  
 KW cardiomyopathy; neurological disease; cancer; AIDS; diagnosis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH CDS 63..3218  
 FT /\*tag= a  
 FT /product= "ataxin-2 like protein"  
 XX  
 PN W09936527-A1.

XX  
 PD 22-JUL-1999.  
 XX  
 PF 19-JAN-1998; 98MO-CN00009.  
 XX  
 PR 19-JAN-1998; 98MO-CN00009.  
 XX  
 PA (UYHU-) UNIV HUMAN MEDICAL.  
 XX  
 PI Deng H, Liu C, Wang D, Xia J;  
 XX  
 DR WPI: 1999-458463/38.  
 DR P-PSDB: AAY29321.  
 XX  
 PS ataxin-2 like protein, and related polynucleotides, useful in  
 PT treatment and diagnosis of ataxia,  
 PT  
 XX  
 PS Claim 4; Page 21-23; 34pp; English.  
 XX  
 CC The present sequence represents a human ataxin-2 like gene. Ataxin-2  
 CC like polynucleotides (1), vectors containing (1) and recombinant host  
 CC cells are useful for recombinant production of ataxin-2 like protein.  
 CC Both (1), ataxin-2 like protein and antibodies against ataxin-2 like  
 CC protein are useful as research reagents, for screening assays and in  
 CC diagnostic assays. Antagonists and agonists of ataxin-2 like protein  
 CC can be used to inhibit or enhance, respectively, the activity of  
 CC ataxin-2 like protein or expression of (1). Anti-ataxin-2 like protein  
 CC antibodies and ataxin-2 like protein or its fragments can be used  
 CC in vaccines. In particular, the proteins, antibodies, agonists and  
 CC antagonists can be used for treating, e.g. ataxia, cardiomyopathy,  
 CC deafness, neurological disease, cancer and AIDS, related to both an  
 CC excess and insufficient amounts of ataxin-2 like.  
 XX  
 SQ Sequence 4674 BP; 971 A; 1566 C; 1244 G; 893 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 4674;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caaggtgcaggtcagc 25  
 |||||  
 DB 3741 CAGGTGCAGGTGACG 3727

RESULT 9  
 AAC00048/c  
 ID AAC00048 standard; cDNA; 327 BP.  
 XX  
 AC AAC00048;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 46.  
 XX  
 KM Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 XX 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI: 2000-500381/45.

DR P-PSDB; AAC00042.  
 XX  
 CC New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 1; SEQ ID 46; 71bp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.  
 CC  
 SQ Sequence 327 BP; 70 A; 102 C; 75 G; 69 T; 11 other;  
 XX  
 XX  
 Query Match 56.0%; Score 14; DB 21; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 12 aggtgcaggtcagc 25  
 DB 144 AGTGCAGGTGACG 131  
 |||||||  
 RESULT 10  
 AAF64981/c  
 ID AAF64981 standard; cDNA; 374 BP.  
 XX  
 AC AAF64981;  
 XX  
 DT 09-APR-2001 (first entry)  
 XX  
 DE Novel human polynucleotide, SEQ ID NO: 737.  
 XX  
 KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;  
 KM breast cancer; lung cancer; cancer detection; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200102568-A2.  
 XX  
 PD 11-JAN-2001.  
 XX  
 PF 30-JUN-2000; 2000WO-US18374.  
 XX  
 PR 02-JUL-1999; 99US-0142310.  
 PR 02-JUL-1999; 99US-0142311.  
 XX  
 PA (CHIR) CHIRON CORP.  
 PA (HSE-) HYSD INC.  
 XX  
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;  
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;  
 PI Chtenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;  
 PI Kita D, Garcia V, Jones LW, Strache-Grain B;  
 DR WPI: 2001-091805/10.  
 XX  
 PT Library of polynucleotides for diagnosing a cancerous state of a  
 PT mammalian cell and detecting cancer, particularly of the colon or  
 PT prostate, comprises 351 human polynucleotide sequences -  
 XX

PS Claim 9; Page 650; 1046bp; English.  
 XX  
 CC The present sequence is one of 351 sequences in a library of human  
 CC polynucleotides. The library is used to detect differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell and can  
 CC detect colon, prostate, breast and lung cancer. The library can be used  
 CC to produce probes for detection of mRNA and to produce additional copies  
 CC of the polynucleotides. The probes can be used for chromosome mapping of  
 CC the polynucleotide and for detection of transcription levels. Ribozymes  
 CC or antisense oligonucleotides can be generated. The polynucleotides and  
 CC their gene products are used as genetic or biochemical markers (e.g. in  
 CC blood or tissues) that will detect the earliest changes along the  
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
 CC preventive interventions. The polynucleotides, polypeptides and  
 CC antibodies against them can be used in pharmaceutical compositions to  
 CC treat the cancers and proliferative disorders such as neoplasia,  
 CC dysplasia and hyperplasia.  
 CC  
 SQ Sequence 374 BP; 71 A; 116 C; 118 G; 68 T; 1 other;  
 XX  
 XX  
 Query Match 56.0%; Score 14; DB 22; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 12 aggtgcaggtcagc 25  
 DB 180 AGTGCAGGTGACG 167  
 |||||||  
 RESULT 11  
 AAQ45597/c  
 ID AAQ45597 standard; DNA; 423 BP.  
 XX  
 AC AAQ45597;  
 XX  
 DT 04-DEC-1993 (first entry)  
 XX  
 DE Sequence encoding the VH of antibody B17X2.  
 XX  
 DE Variable heavy antibody chain; human subgroup 4 germline; ss.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 7..423  
 FT /tag= a  
 XX  
 PN WO9312231-A.  
 XX  
 PD 24-JUN-1993.  
 XX  
 PF 13-DEC-1991; 91WO-AU00583.  
 XX  
 PR 13-DEC-1991; 91WO-AU00583.  
 XX  
 PA (DOWC) DOW CHEM AUSTRALIA LTD.  
 XX  
 PI Johnson KS, Mezes PS, Richard RA;  
 DR WPI: 1993-214173/26.  
 DR P-PSDB; AAR8315.  
 XX  
 PT New composite antibody binding to tumour associated TAG-72  
 PT antigen - includes light chain variable region from human  
 PT subgroup 4 germline gene, useful, opt. as conjugate, for  
 PT diagnosis or treatment of cancer  
 XX  
 PS Disclosure; Figure 4; 150bp; English.  
 XX  
 CC Cell line B17X2 expresses an antibody utilizing a variable light  
 CC chain encoded by a gene derived from Humd VL and a variable heavy  
 CC chain which makes a stable VL and VH combination.  
 CC

XX Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;

Query Match 56.0%; Score 14; DB 14; Length 423;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcaggtcagc 25  
|||||  
DB 133 AGTGCAGGTCAGC 120

RESULT 12  
AA239427/C  
ID AA239427 standard; DNA; 423 BP.

XX AA239427;

XX 29-FEB-2000 (first entry)

XX Antibody B17x2 VH nucleotide sequence.

XX Tumor associated stialylated glycoprotein: TAG-72; cancer antigen;  
KW carcinoma lesion; diagnostic; cancer; antibody; human; B17x2;  
KW anti-mouse antibody hypersensitivity reaction; ss.

XX Homo sapiens.

XX US5976845-A.

XX 02-NOV-1999.

XX 07-JUN-1995; 95US-0487743.

XX 16-JUN-1994; 94US-0261354.

XX 19-APR-1990; 90US-0510697.

XX 20-OCT-1992; 92US-0964536.

XX (DOWC ) DOW CHEM CO.

XX Johnson KS, Richard RA, Mezes PS;

XX WPI, 1999-619651/53.

XX P-PSDB; AAV57179.

XX Production of humanized anti-TAG-72 antibodies, used for the detection,  
PT in vivo imaging and treatment of cancers -  
XX

XX Disclosure: Fig 4A-B; 85pp; English.

XX The invention relates to producing humanized anti-tumor associated  
CC stialylated glycoprotein (TAG-72) antibodies (anti-TAG). The antibodies  
CC have binding specificity for the cancer antigen TAG-72. These antibodies  
CC have variable regions with VL segments derived from human subgroup IV  
CC germ-line gene and a VH segment (encoded by the VHA1phatag germ-line gene)  
CC which is capable of combining with the VL to form a three dimensional  
CC structure having the ability to bind TAG-72. They can be used for the in  
CC vivo detection of carcinoma lesions. They can also be used for in vitro  
CC diagnostics. They can also be modified with therapeutic agents e.g. a  
CC radionuclide, drug, biological response modifier, toxin or another  
CC antibody for the treatment of cancers. The humanized anti-TAG-72  
CC antibodies can reduce harmful anti-mouse antibody hypersensitivity  
CC reactions.

XX Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;

Query Match 56.0%; Score 14; DB 20; Length 423;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcaggtcagc 25

DB 133 AGTGCAGGTCAGC 120  
|||||

RESULT 13  
AA23965/C  
ID AA23965 standard; DNA; 423 BP.

XX AA23965;

XX 09-FEB-2000 (first entry)

XX Human B17X2 antibody VH segment DNA.

XX Human; antibody; humanized; anti-tumor; stialylated glycoprotein antibody;  
KW TAG-72; cytostatic; cancer antigen; detection; carcinoma lesion;  
KW diagnostic; treatment; ss.

XX Homo sapiens.

XX US5976531-A.

XX 02-NOV-1999.

XX 16-JUN-1994; 94US-0261354.

XX 19-APR-1990; 90US-0510697.

XX 20-OCT-1992; 92US-0964536.

XX (DOWC ) DOW CHEM CO.

XX Johnson KS, Mezes PS, Richard RA;

XX WPI, 1999-632731/54.

XX P-PSDB; AAV50688.

XX New humanized anti-TAG-72 antibodies, used for the detection, in vivo  
PT imaging and treatment of cancers -  
XX

XX Disclosure: Figure 4A-B; 83pp; English.

XX This invention describes novel humanized anti-tumor associated  
CC stialylated glycoprotein antibodies (TAG-72) which have cytostatic  
CC activity. The antibodies have binding specificity for the cancer antigen  
CC TAG-72. They can be used for the in vivo detection of carcinoma lesions.  
CC They can also be used for in vitro diagnostics. They can also be modified  
CC with therapeutic agents e.g. a radionuclide, drug, biological response  
CC modifier, toxin or another antibody for the treatment of cancers. The  
CC humanized anti-TAG-72 antibodies can reduce harmful anti-mouse antibody  
CC hypersensitivity reactions.

XX Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;

Query Match 56.0%; Score 14; DB 20; Length 423;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcaggtcagc 25  
|||||  
DB 133 AGTGCAGGTCAGC 120

RESULT 14  
AAT80181/C  
ID AAT80181 standard; DNA; 444 BP.

XX AAT80181;

XX 02-APR-1998 (first entry)

XX Monoclonal antibody 105AD7 heavy chain variable region DNA sequence.

XX

KW Monoclonal antibody 105AD7; complementarity determining region;  
 KW CDR; human; immune response; treatment; cancer; ss.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 XX CDS 1..444  
 XX FT /\*tag= a  
 XX FT /transl\_except= (pos: 157..159, aa: Gly)  
 XX FT /transl\_except= (pos: 160..162, aa: Val)  
 XX FT /product= "105AD7 heavy chain variable region"  
 XX  
 XX MO9732021-A1.  
 XX  
 XX PD 04-SEP-1997.  
 XX  
 XX PF 28-FEB-1997; 97WO-GB00591.  
 XX  
 XX PR 29-FEB-1996; 96GB-0004321.  
 XX PR 28-FEB-1996; 96GB-0004177.  
 XX  
 XX PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY LTD.  
 XX  
 XX PI Durrant LG, Robins RA, Spendlove I;  
 XX  
 XX DR WPI: 1997-448689/41.  
 XX DR P-PSDB: AAW26239.  
 XX  
 XX PT Nucleic acid encoding peptide derived from antibody 105AD7 - used to  
 XX PT induce an immune response for treatment and prevention of cancer  
 XX  
 XX PS Claim 1; Fig 1A; 32pp; English.  
 XX  
 XX CC This is the DNA sequence of the monoclonal antibody 105AD7 heavy chain  
 XX CC variable region. 105AD7 is a human monoclonal anti-idiotypic antibody  
 XX CC that mimics T cell epitopes on the tumour associated antigen gp72 and  
 XX CC has been used for therapeutic vaccination. A peptide derived from this  
 XX CC antibody contains the 105AD7 heavy chain variable region and a 105AD7  
 XX CC kappa chain. The complementarity determining regions (CDR) of this  
 XX CC peptide and its fragments can stimulate immune responses. The peptide  
 XX CC can be used for screening for its functional immune responses. The peptide  
 XX CC stimulates an immune response to a tumour antigen, for the treatment or  
 XX CC prevention of tumours. Transfected host cells are used to produce this  
 XX CC peptide and the vectors can be used to generate the peptide in vivo for  
 XX CC stimulation of an immune response. The peptide contains promiscuous  
 XX CC helper epitopes which stimulate a response against tumour antigens other  
 XX CC than gp72. These epitopes may help a cytotoxic T lymphocyte response to  
 XX CC any co-injected antigen.  
 XX  
 XX SQ Sequence 444 BP; 100 A; 134 C; 107 G; 103 T; 0 other;  
 XX  
 XX Query Match 56.0%; Score 14; DB 18; Length 444;  
 XX Best Local Similarity 100.0%; Pred. No. 74;  
 XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX QY 12 aggtcaggtcagc 25  
 XX ||||||||||||  
 XX DB 121 AGGTCCAGGTGACG 108  
 XX  
 XX RESULT 15  
 XX AA220406/C  
 XX ID AA220406 standard; cDNA; 447 BP.  
 XX  
 XX AC AA220406;  
 XX  
 XX DT 19-NOV-1999 (first entry)  
 XX  
 XX XX IgG antibody 2.4.4 heavy chain coding sequence.  
 XX DE  
 XX XX Antibody; CD147; IgM; ABX-CBL; activated T-cell killing; leukaemia;  
 XX KW

KW activated B-cell; monocyte; graft versus host disease; therapy; cancer;  
 KW organ transplant rejection disease; lymphoma; pancreatic disease;  
 KW autoimmune disease; inflammatory disease; arthritis; binding site; ss.  
 OS Homo sapiens.  
 XX  
 XX XX Key Location/Qualifiers  
 XX CDS 1..447  
 XX FT /\*tag= a  
 XX FT /note= "no stop codon given"  
 XX  
 XX PN MO9945031-A2.  
 XX  
 XX PD 10-SEP-1999.  
 XX  
 XX PF 03-MAR-1999; 99WO-US04583.  
 XX  
 XX PR 03-MAR-1998; 98US-0034607.  
 XX PR 03-FEB-1999; 99US-0244253.  
 XX  
 XX PA (ABGE-) ABGENIX INC.  
 XX  
 XX PI Davis CG, Blacher RW, Corvatan JR, Culwell AR, Green LT, Hales J;  
 XX PI Havrilla N, Ivanov VE, Lipani JA, Liu Q, Weber RF, Yang X;  
 XX  
 XX DR WPI: 1999-540816/45.  
 XX DR P-PSDB: AAY34305.  
 XX  
 XX PT New monoclonal antibody, used for treating e.g. graft versus host  
 XX PT disease, cancers, autoimmune diseases and inflammatory diseases -  
 XX  
 XX PS Disclosure; Fig 30; 245pp; English.  
 XX  
 XX CC This sequence encodes the heavy chain of an antibody of the  
 XX CC invention. The antibody is a monoclonal antibody (Mab) with an isotype  
 XX CC that fixes complement and a variable region that binds to the epitope on  
 XX CC CD147 bound by the IgM Mab ABX-CBL, providing that the antibody is not  
 XX CC CBL. The Mab can selectively kill activated T-cells, activated B-cells  
 XX CC or resting or activated monocytes. The products and methods can be used  
 XX CC for treating diseases involving activated T-cells or B-cells or  
 XX CC monocytes, e.g. graft versus host disease (GVHD), organ transplant  
 XX CC rejection diseases (e.g. renal transplant, ocular transplant), cancers  
 XX CC (e.g. cancers of the blood (e.g. leukaemia's and lymphomas) and  
 XX CC pancreatic), autoimmune diseases (e.g. lupus), and inflammatory diseases  
 XX CC (e.g. arthritis).  
 XX  
 XX SQ Sequence 447 BP; 96 A; 150 C; 106 G; 95 T; 0 other;  
 XX  
 XX Query Match 56.0%; Score 14; DB 20; Length 447;  
 XX Best Local Similarity 100.0%; Pred. No. 74;  
 XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX QY 12 aggtcaggtcagc 25  
 XX ||||||||||||  
 XX DB 34 AGGTCCAGGTGACG 21  
 XX  
 XX RESULT 16  
 XX AA078943/C  
 XX ID AA078943 standard; DNA; 613 BP.  
 XX  
 XX AC AA078943;  
 XX  
 XX DT 07-AUG-1995 (first entry)  
 XX  
 XX XX Human immunoglobulin variable heavy chain #5.  
 XX DE  
 XX XX Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;  
 XX KW cosmid; placenta; vector; pUB81; E.coli; mammalian; ds.  
 XX  
 XX OS Homo sapiens.  
 XX KW

```

FH Key Location/Qualifiers
FT CDS 71..513
FT /tag= a
FT /product= human immunoglobulin variable heavy chain
FT Intron 117..202
FT /tag= b
FT misc_signal 514..516
FT /tag= c
FT /note= "miscellaneous signal, does not conform to
FT terminator or splice site sequence"
FT
FT WO9426895-A.
FT
FT 24-NOV-1994.
FT
FT 10-MAY-1993; 93WO-JP00603.
FT
FT 10-MAY-1993; 93WO-JP00603.
FT
FT (NISB ) JAPAN TOBACCO INC.
FT
FT Honjo T, Matsuda F;
FT WPI; 1995-006791/01.
FT P-PSDB; AAR66299.
FT
FT DNA fragment comprising human immunoglobulin Vh genes - for the
FT production of human immunoglobulin in mammalian hosts
FT
FT Disclosure: Page 36-37; 130pp; Japanese.
FT
XX
XX A series of genes (AA078939-79002) encoding human immunoglobulin
XX variable heavy chains. The genes were isolated and cloned from a series
XX of cosmid constructs: Y202; Y103; Y21; Y6; Y24; 3-31; M84; M118 and M131,
XX by PCR amplification using primers AA078917-38. The genes are subdivided
XX into 5 families of Vh genes. The fragments cover a region of 800 kb.
XX The DNA fragments were isolated from high molecular weight DNA from
XX human placenta. The DNA was partially digested with Taqi restriction
XX enzyme. The fragments were separated by gel electrophoresis and 35-45 kb
XX fractions were collected. The fragments were ligated with ClaI-digested
XX cosmid vector pUB81. The ligation products were in vitro packed and
XX infected into E.coli 490A. The fragments were then subcloned by colony
XX hybridisation. The Vh genes and the DNA fragments encoding them are
XX useful in producing human immunoglobulin in mammalian hosts.
XX
SQ Sequence 613 BP; 139 A; 194 C; 139 G; 141 T; 0 other;

Query Match 56.0%; Score 14; DB 16; Length 613;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcaggtcagc 25
   |||||||
DB 283 AGGTGACAGTCTCAGC 270

RESULT 17
AA078948/c
ID AA078948 standard; DNA; 630 BP.
XX
XX AA078948;
AC
XX
XX 01-AUG-1995 (first entry)
DE
XX Human immunoglobulin Vh gene #10.
XX
XX Primer: PCR; amplify: human; immunoglobulin; variable; heavy chain;
XX cosmid; placenta; vector; pUB81; E.coli; mammalian; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers

```

```

FT CDS 71..495
FT /tag= a
FT /product= human immunoglobulin variable heavy chain
FT Intron 114..199
FT /tag= b
FT misc_signal 322..324
FT /tag= c
FT /transl_except= unused termination codon
FT misc_signal 373..375
FT /tag= d
FT /transl_except= unused termination codon
FT misc_signal 496..498
FT /tag= e
FT /note= "miscellaneous signal, does not conform to
FT termination or splice site sequence"
FT
FT WO9426895-A.
FT
FT 24-NOV-1994.
FT
FT 10-MAY-1993; 93WO-JP00603.
FT
FT 10-MAY-1993; 93WO-JP00603.
FT
FT (NISB ) JAPAN TOBACCO INC.
FT
FT Honjo T, Matsuda F;
FT WPI; 1995-006791/01.
FT P-PSDB; AAR66304.
FT
FT DNA fragment comprising human immunoglobulin Vh genes - for the
FT production of human immunoglobulin in mammalian hosts
FT
FT Claim 19; Page 43-44; 130pp; Japanese.
FT
XX
XX A series of genes (AA078939-79002) encoding human immunoglobulin
XX variable heavy chains. The genes were isolated and cloned from a series
XX of cosmid constructs: Y202; Y103; Y21; Y6; Y24; 3-31; M84; M118 and M131,
XX by PCR amplification using primers AA078917-38. The genes are subdivided
XX into 5 families of Vh genes. The fragments cover a region of 800 kb.
XX The DNA fragments were isolated from high molecular weight DNA from
XX human placenta. The DNA was partially digested with Taqi restriction
XX enzyme. The fragments were separated by gel electrophoresis and 35-45 kb
XX fractions were collected. The fragments were ligated with ClaI-digested
XX cosmid vector pUB81. The ligation products were in vitro packed and
XX infected into E.coli 490A. The fragments were then subcloned by colony
XX hybridisation. The Vh genes and the DNA fragments encoding them are
XX useful in producing human immunoglobulin in mammalian hosts.
XX
SQ Sequence 630 BP; 141 A; 179 C; 148 G; 162 T; 0 other;

Query Match 56.0%; Score 14; DB 16; Length 630;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcaggtcagc 25
   |||||||
DB 280 AGGTGACAGTCTCAGC 267

RESULT 18
AA015905/c
ID AA015905 standard; cDNA; 981 BP.
XX
XX AA015905;
AC
XX
XX 12-JUN-2000 (first entry)
DE
XX Human protein clone HP10568 coding sequence.
XX
XX Human protein; hydrophobic domain; nutritional source; haematopoiesis;
XX
XX Key Location/Qualifiers

```

KM cytokine production; cell proliferation; cell differentiation;  
 KM immune deficiency; infectious disease; autoimmune disorder; asthma;  
 KM multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KM allergic reaction; osteoporosis; osteoarthritis; periodontal disease;  
 KM nervous system disorder; Alzheimer's disease; Parkinson's disease;  
 KM Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;  
 KM systemic cytokine damage; tissue differentiation; contraceptive; stroke;  
 KM coagulation disorder; myocardial infarction; inflammatory condition;  
 KM septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;  
 KM nephritis; therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200005367-A2.  
 PN  
 PD 03-FEB-2000.  
 XX  
 XX 22-JUL-1999; 99WO-JP03929.  
 PE  
 XX 24-JUL-1998; 98JP-0208820.  
 PR 07-AUG-1998; 98JP-0224105.  
 PR 25-AUG-1998; 98JP-0238116.  
 PR 09-SEP-1998; 98JP-0254736.  
 PR 29-SEP-1998; 98JP-0275505.  
 XX  
 PA (SAGA ) SAGAMI CHEM RES CENT.  
 PA (PROT-) PROTEGENE INC.  
 PI Kato S, Kimura T;  
 PI  
 XX WPI: 2000-182694/16.  
 DR P-PSDB: AAY94857.  
 XX  
 PT Novel human proteins having hydrophobic domains useful for treating  
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,  
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -  
 XX  
 PS Claim 3; Page 189; 351pp; English.

XX  
 SQ Sequence 981 BP; 194 A; 299 C; 275 G; 213 T; 0 other;  
 XX  
 Query Match 56.0%; Score 14; DB 21; Length 981;  
 Best local Similarity 100.0%; Pred. No. 69;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 12 aggtgcaggtcagc 25  
 |||||  
 Db 139 AGGTGCAGGTCAAGC 126  
 RESULT 19  
 AAA15915/c  
 ID AAA15915 standard; cDNA; 1100 BP.  
 XX  
 AC AAA15915;  
 XX  
 DT 12-JUN-2000 (first entry)  
 XX  
 DE Human protein clone HP10568 full length coding sequence.  
 XX  
 KM Human protein; hydrophobic domain; nutritional source; haematopoiesis;  
 KM cytokine production; cell proliferation; cell differentiation;  
 KM immune deficiency; infectious disease; autoimmune disorder; asthma;  
 KM multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KM allergic reaction; osteoporosis; osteoarthritis; periodontal disease;  
 KM nervous system disorder; Alzheimer's disease; Parkinson's disease;  
 KM Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;  
 KM systemic cytokine damage; tissue differentiation; contraceptive; stroke;  
 KM coagulation disorder; myocardial infarction; inflammatory condition;  
 KM septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;  
 KM nephritis; therapy; ss.  
 KM  
 OS Homo sapiens.  
 XX  
 PN WO200005367-A2.  
 PD 03-FEB-2000.  
 XX  
 PF 22-JUL-1999; 99WO-JP03929.  
 XX  
 PR 24-JUL-1998; 98JP-0208820.  
 PR 07-AUG-1998; 98JP-0224105.  
 PR 25-AUG-1998; 98JP-0238116.  
 PR 09-SEP-1998; 98JP-0254736.  
 PR 29-SEP-1998; 98JP-0275505.  
 XX  
 PA (SAGA ) SAGAMI CHEM RES CENT.  
 PA (PROT-) PROTEGENE INC.  
 PI Kato S, Kimura T;  
 PI  
 XX WPI: 2000-182694/16.  
 DR P-PSDB: AAY94857.  
 XX  
 PT Novel human proteins having hydrophobic domains useful for treating  
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,  
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -  
 XX  
 PS Claim 4; Page 203-205; 351pp; English.

This sequence encodes a human protein of the invention, which has hydrophobic domains. The DNA sequences can be used as a probe or as a genetic marker. The protein can also be used as a marker, and to identify potential genetic disorders. The DNA and protein can also be used as nutritional sources or supplements. The protein exhibits cytokine, cell proliferation, cell differentiation activities and induces production of other cytokines in certain cell populations. The protein also exhibits immune stimulating or immune suppressing activity. It can be used in the treatment of various immune deficiencies and disorders, and to treat infectious diseases caused by viral, bacterial, fungal or other

infections. The protein is also used for treating autoimmune disorders such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid arthritis. It is also useful in the treatment of allergic reactions and conditions such as asthma, and in immune suppression after organ transplantation. The protein is useful in regulation of haematopoiesis and consequently in the treatment of myeloid or lymphoid cell deficiencies. It is also used in compositions for tissue growth or regeneration. The protein is also used in the treatment of osteoporosis or osteoarthritis and in the treatment of periodontal disease and other tooth repair processes. The protein is used in the treatment of nervous system disorders such as Alzheimer's disease, Parkinson's disease, and Huntington's disease. They are useful for protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage. They are also used for promoting or inhibiting tissue differentiation. They are also used as contraceptives since they exhibit activin or inhibin related activities and as a fertility inducing therapeutic. They are used for treating various coagulation disorders and in treatment and prevention of conditions resulting from coagulation activities e.g. myocardial infarction or stroke. They also acts as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. They are used to treat inflammatory conditions such as septic shock, sepsis, ischaemia reperfusion injury, arthritis, and nephritis. They can be used to prevent tumours.

Sequence 1100 BP, 221 A; 333 C; 314 G; 232 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcagtcagc 25  
|||||  
DB 195 AGCTGCAGTCAGC 182

RESULT 20  
AAA39064/C  
ID AAA39064 standard; cDNA: 1130 BP.  
XX  
AC AAA39064;  
XX  
DT 30-AUG-2000 (first entry)  
XX  
DE Human secreted protein gene 13 SEQ ID NO:23.  
XX  
KW Human: secreted protein; cytosolic; anti-proliferative; vulnary;  
KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;  
KW hyperproliferative disorder; infectious disease; tissue regeneration;  
KW screening; food additive; preservative; wound healing;  
KW hyper-vascular disease; chromosome 11; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200017222-A1.  
XX  
PD 30-MAR-2000.  
XX  
PF 22-SEP-1999; 99WO-US22012.  
XX  
PR 23-SEP-1998; 98US-0101546.  
PR 02-OCT-1998; 98US-0102895.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;  
PI Komatsoulis G, Endress GA, Soppet DR;  
XX  
DR WPI: 2000-283538/24.  
DR P-PSDB: AAB08903.  
XX  
PT Human secreted proteins and coding sequences useful in diagnostic and

therapeutic methods for disorders such as immune system or proliferative disorders, related to the proteins  
Claim 1; Page 329-330; 416pp; English.

The polynucleotide sequences given in AAA39052 to AAA39088 encode the human secreted proteins given in AAB08891 to AAB08984. The human secreted proteins can have activities based on the tissues and cells they are expressed in. Examples of the activities are: cytosolic; and vulnary. The anti-proliferative; immunosuppressive; antibacterial; and vulnary. The secreted proteins and their related polynucleotide sequences are useful for diagnostic and therapeutic methods useful for diagnosing and treating disorders related to the secreted proteins. The proteins, and polynucleotide sequences may be useful for treating disorders of the immune system, hyperproliferative disorders, infectious disease, regeneration of tissues, for chemotaxis and for screening molecules that bind to the proteins. The proteins or polynucleotide sequences may be used as food additives or preservatives, to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional components. Agonists or antagonists of the proteins may be used to prevent scar tissue growth during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051 and AAB08890 are sequences used in the exemplification of the present invention.

Sequence 1130 BP; 246 A; 334 C; 317 G; 233 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1130;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcagtcagc 25  
|||||  
DB 201 AGCTGCAGTCAGC 188

RESULT 21  
AAA39087/C  
ID AAA39087 standard; cDNA: 1135 BP.  
XX  
AC AAA39087;  
XX  
DT 30-AUG-2000 (first entry)  
XX  
DE Human secreted protein gene 13 SEQ ID NO:46.  
XX  
KW Human: secreted protein; cytosolic; anti-proliferative; vulnary;  
KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;  
KW hyperproliferative disorder; infectious disease; tissue regeneration;  
KW screening; food additive; preservative; wound healing;  
KW hyper-vascular disease; chromosome 11; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200017222-A1.  
XX  
PD 30-MAR-2000.  
XX  
PF 22-SEP-1999; 99WO-US22012.  
XX  
PR 23-SEP-1998; 98US-0101546.  
PR 02-OCT-1998; 98US-0102895.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;  
PI Komatsoulis G, Endress GA, Soppet DR;  
XX  
DR WPI: 2000-283538/24.  
DR P-PSDB: AAB08926.  
XX  
PT Human secreted proteins and coding sequences useful in diagnostic and

PT therapeutic methods for disorders such as immune system or  
 PT proliferative disorders, related to the proteins  
 PS Claim 1; Page 347-348; 416pp; English.  
 XX  
 CC The polynucleotide sequences given in AAA39052 to AAA39088 encode the  
 CC human secreted proteins given in AAB08891 to AAB08984. The human secreted  
 CC proteins can have activities based on the tissues and cells they are  
 CC expressed in. Examples of the activities are: cytostatic;  
 CC anti-proliferative; immunosuppressive; antibacterial; and vulnerary. The  
 CC secreted proteins and their related polynucleotide sequences are useful  
 CC for diagnostic and therapeutic methods useful for diagnosing and treating  
 CC disorders related to the secreted proteins. The proteins, and  
 CC polynucleotide sequences may be useful for treating disorders of the  
 CC immune system, hyperproliferative disorders, infectious disease,  
 CC regeneration of tissues, for chemotaxis and for screening molecules that  
 CC bind to the proteins. The proteins or polynucleotide sequences may be  
 CC used as food additives or preservatives, to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, co-factors or other nutritional components. Agonists or  
 CC antagonists of the proteins may be used to prevent scar tissue growth  
 CC during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051  
 CC and AAB08890 are sequences used in the exemplification of the present  
 CC invention.  
 XX  
 SO Sequence 1135 BP; 243 A; 333 C; 323 G; 234 T; 2 other;  
 QY  
 Query Match 56.0%; Score 14; DB 21; Length 1135;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 12 aggtgcaggtgcagc 25  
 188 AGGTGCAGGTGCAGC 175  
 RESULT 22  
 AAA29322  
 ID AAA29322 standard; cDNA; 1160 BP.  
 XX  
 AC AAA29322;  
 XX  
 DT 26-SEP-2000 (first entry)  
 XX  
 DE Glycine max isoflavone O-methyltransferase clone srl.pk0015.b4.  
 XX  
 KW Isoflavone O-methyltransferase; soybean; degradation; daidzein;  
 KW Flavonoid; flower colour; growth; pollination; irradiation; ss.  
 XX  
 OS Glycine max.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 11..1087  
 FT /tag= a  
 FT /product= Isoflavone\_O-methyltransferase  
 XX  
 PN WO200037656-A2.  
 XX  
 PD 29-JUN-2000.  
 XX  
 PF 20-DEC-1999; 99WO-US30338.  
 XX  
 PR 21-DEC-1998; 98US-0113190.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E. I.  
 XX  
 PI Famodu OO, Mcgonigle B, Odell JT, Fader GM, Falco SC;  
 DR WPI: 2000-442680/38.  
 DR P-PSDB: AAY96584.  
 XX  
 PT New polynucleotide encoding flavonoid biosynthetic enzymes for creating

PT transgenic plants and for immunological screening of cDNA libraries  
 XX  
 PS Claim 2; Page 32; 39pp; English.  
 XX  
 CC AAA29321-25 encode isoflavone O-methyltransferases isolated from various  
 CC soybean (Glycine max) tissues. cDNA libraries se4, srl, srlc and sllc  
 CC were prepared from soybean embryo (19 days after flowering), root, 8-day  
 CC old root and seed (25 days after flowering).  
 CC Isoflavone O-methyltransferase catalyses the first step in degradation  
 CC of daidzein. Suppression of this enzyme will yield higher concentrations  
 CC of this beneficial isoflavone in, e.g. soybean seed. Flavonoids serve as  
 CC co-pigments in flower colour, stimulate pollen tube growth, attract  
 CC pollinators, act as feeding deterrents and protect against UV  
 CC irradiation in fruits and seeds. The cDNA and proteins can be used to  
 CC isolate homologues, for immunological screening and for positive  
 CC selection methods.  
 XX  
 SO Sequence 1160 BP; 361 A; 197 C; 254 G; 348 T; 0 other;  
 QY  
 Query Match 56.0%; Score 14; DB 21; Length 1160;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 222 ctctcaggtgcagc 235  
 6 ctctcaggtgcagc 19  
 222 ctctcaggtgcagc 235  
 RESULT 23  
 AAA91019/c  
 ID AAA91019 standard; DNA; 1161 BP.  
 XX  
 AC AAA91019;  
 XX  
 DT 05-APR-2001 (first entry)  
 XX  
 DE Human secreted protein PRO7154 coding sequence.  
 XX  
 KW Secreted protein; human; PRO protein; neoplastic cell growth; tumour;  
 KW proliferation; leukemia; lymphoid malignancy; inflammatory disorder;  
 KW angiogenic disorder; immunologic disorder; PRO7154; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 102..1085  
 FT /tag= a  
 FT /product= PRO7154  
 XX  
 PN WO200075317-A2.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PF 15-MAY-2000; 2000WO-US13358.  
 XX  
 PR 09-JUN-1999; 99US-0138385.  
 PR 20-JUL-1999; 99US-0144790.  
 PR 03-AUG-1999; 99US-0146843.  
 PR 10-AUG-1999; 99US-0148188.  
 PR 17-AUG-1999; 99US-0149320.  
 PR 17-AUG-1999; 99US-0149327.  
 PR 17-AUG-1999; 99US-0149396.  
 PR 20-AUG-1999; 99US-0150114.  
 PR 31-AUG-1999; 99US-0151700.  
 PR 31-AUG-1999; 99US-0151734.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Botstein DA, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;  
 DR WPI: 2001-071075/08.  
 DR P-PSDB: AAY97585.  
 XX  
 PT



XX Antibodies against PRO polypeptides, useful for diagnosing and treating  
 PT tumours are associated with gene amplification, neoplastic cell growth  
 PT and proliferation in mammals -  
 XX  
 PS Claim 50; Fig 11; 143pp; English.

XX This sequence encodes a human PRO protein of the invention. The PRO  
 CC proteins are secreted proteins. Antagonists or antibodies of PRO  
 CC polypeptides are useful for diagnosing and treating tumours are  
 CC associated with gene amplification, neoplastic cell growth and  
 CC proliferation in mammals, and those conditions characterised by  
 CC overexpression and/or activation of the amplified genes. Such conditions  
 CC include benign or malignant tumours (e.g. renal, liver, kidney, bladder,  
 CC breast, gastric, ovarian, colorectal, prostate, pancreatic, lung, vulval,  
 CC thyroid, hepatic carcinomas, sarcomas, glioblastomas and various head and  
 CC neck tumours); leukaemias and lymphoid malignancies; neuronal, glial,  
 CC astrocytal, hypothalamic, and other glandular, macrophageal, epithelial,  
 CC stromal and blastocoele disorders; and inflammatory, angiogenic and  
 CC immunologic disorders. These may further be used to qualitatively or  
 CC quantitatively detect the expression of proteins encoded by the  
 CC amplified genes, and in tumour diagnostics or prognostics. The PRO  
 CC polypeptide or its antagonist may be used for the preparation of a  
 CC medicament in the treatment of a condition, which is responsive to the  
 CC PRO polypeptide, its antagonist or anti-PRO antibody.

XX Sequence 1161 BP; 235 A; 356 C; 327 G; 242 T; 1 other;

SO

Query Match 56.0%; Score 14; DB 22; Length 1161;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcaggtcagc 25  
 |||||  
 DB 240 AGGTGCAGGTCTCAGC 227

RESULT 24  
 AA298136/c  
 ID AA298136 standard; cDNA; 1173 BP.

XX  
 AC AA298136;

DT 11-MAY-2000 (first entry)

XX Human signal peptide containing protein HSP-28 cDNA SEQ ID NO:162.

DE  
 XX  
 XX Human: signal peptide-containing protein; HSP: diagnosis; cancer;  
 KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;  
 KW antimicrobial; nocotropic; neuroprotective; cardiovascular; hepatotropic;  
 KW antistimatic; gene therapy; cell proliferation; neurological disorder;  
 KW reproductive disorder; developmental disorder; arteriosclerosis;  
 KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;  
 KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;  
 KW Parkinson's disease; Huntington's diseases; ovulatory defect;  
 KW muscular dystrophy; ss.

XX  
 XX Homo sapiens.

OS  
 XX  
 PN WO200000610-A2.

XX 06-JAN-2000.

PD  
 XX  
 PF 25-JUN-1999; 99WO-US14484.

XX  
 XX 26-JUN-1998; 98US-0090762.  
 PR 31-JUN-1998; 98US-0094983.  
 PR 01-OCT-1998; 98US-0102686.  
 PR 11-DEC-1998; 98US-0112129.

XX  
 XX (INCY-) INCYTE PHARM INC.

PI Lal P. Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;  
 PI Akerbloom IE, Au-Young J, Yve H, Patterson C, Reddy R, Hillman JL;  
 PI Bandman O;  
 XX  
 DR WPI: 2000-160673/14.  
 DR P-PSDB: AAY87251.

XX  
 XX  
 PT New human signal peptide-containing proteins useful in treatment,  
 PT prevention and diagnosis of e.g. cancer, inflammation and  
 PT cardiovascular disease -  
 XX  
 PS Claim 9; Page 268; 327pp; English.

XX AA298109 to AA298242 encode AAY87224 to AAY87257 which represent the  
 CC human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have  
 CC anticancer, anti-inflammatory, antimicrobial, nocotropic, hepatotropic,  
 CC neuroprotective, cardiovascular and antistimatic activities, and can  
 CC be used in gene therapy. HSPs can be used to treat or prevent disorders  
 CC associated with decreased activity or function of HSP. Antagonists of  
 CC HSP are used to treat or prevent disorders associated with increased  
 CC activity or function of HSP. Such diseases include cell proliferation  
 CC (including cancer), inflammation, cardiovascular, neurological,  
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,  
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,  
 CC asthma, Crohn's disease, microbial or other infections, congestive or  
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's  
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP  
 CC nucleic acids can be used for the recombinant production of HSP, for  
 CC detecting HSP in standard hybridisation and amplification assays (for  
 CC diagnosis and monitoring), in gene therapy, as antisense,  
 CC triplex-forming or ribozyme therapeutics, for detecting related sequences  
 CC or genetic variations, and for chromosomal mapping. HSP are also used to  
 CC raise specific antibodies (Ab) and to screen for agonists and  
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or  
 CC monitor, HSP-related diseases (in usual immunoassays), as therapeutic  
 CC antagonists, in competitive drug screens, and for purification of HSP  
 CC from natural sources.

XX Sequence 1173 BP; 236 A; 363 C; 330 G; 244 T; 0 other;

SO

Query Match 56.0%; Score 14; DB 21; Length 1173;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcaggtcagc 25  
 |||||  
 DB 256 AGGTGCAGGTCTCAGC 243

RESULT 25  
 AAC98064/c  
 ID AAC98064 standard; cDNA; 1186 BP.

XX  
 AC AAC98064;

DT 09-MAR-2001 (first entry)

XX Human colon cancer antigen nucleotide sequence SEQ ID NO:74.

DE  
 XX  
 XX Human: colon cancer antigen; diagnosis; detection;  
 KW identification; cytostatic; cardioactive; neuroprotective; vulnerrary;  
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;  
 KW neural disorder; immune system disorder; muscular disorder;  
 KW reproductive disorder; gastrointestinal disorder; renal disorder;  
 KW infectious disease; cardiovascular disorder; ss.

XX  
 XX Homo sapiens.

OS  
 XX  
 PN WO20005351-A1.

XX 21-SEP-2000.



PF 26-MAY-2000; 2000MO-US14842.  
 XX  
 PR 28-MAY-1999; 99US-0136503.  
 XX  
 PA (SUGEN-) SUGEN INC.  
 XX  
 PI Plowman GD, Martinez R, Whyte D, Sudersanam S;  
 XX  
 DR MPI; 2001-032161/04.  
 DR P-PSDB; AAB65598.  
 XX  
 PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and  
 PT treating immune-related diseases and disorders, cardiovascular disease,  
 XX neurodegenerative diseases and/or cancers -  
 XX  
 PS Disclosure; Fig 2; 310pp; English.  
 XX  
 CC The present sequence encodes a novel protein kinase. The nucleic acids  
 CC and the protein kinases they encode may be used in the treatment and  
 CC diagnosis of diseases associated with inappropriate kinase expression  
 CC such as immune-related diseases and disorders, cardiovascular disease,  
 CC neurodegenerative diseases and/or cancers. The nucleic acids and  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC assays. The kinase polypeptides may be used as antigens in the production  
 CC of antibodies or kinase expression and activity. Anti-kinase antibodies  
 CC and kinase antagonists may also be used to down regulate kinase  
 CC expression and activity. Diseases related to kinase expression and  
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune  
 CC disorders, complications of organ transplantation, myocardial infarction,  
 CC immune disorders, cardiomyopathies, strokes, renal failure,  
 CC oxidative stress related disorders, chronic inflammatory bowel disease,  
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,  
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and  
 CC reproductive disorders.  
 XX  
 SQ Sequence 1368 BP; 346 A; 359 C; 398 G; 265 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1368;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcaggtcagc 25  
 |||||||||  
 DB 1022 AGGTGAGGTGAGC 1009

RESULT 28  
 AA296242  
 ID AA296242 standard; DNA; 1634 BP.  
 XX  
 AC AA296242;  
 XX  
 DT 10-APR-2000 (first entry)  
 XX  
 DE S. pneumoniae derived DNA from ORF #70.  
 XX  
 KW Treatment; prevention; disease; diagnosis; gene therapy; screening;  
 XX bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.  
 OS Streptococcus pneumoniae.  
 XX  
 PN MO9806734-A1.  
 PD 19-FEB-1998.  
 XX  
 PF 15-AUG-1997; 97WO-US14436.  
 XX  
 PR 16-AUG-1996; 96US-0024022.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
 DR

PI Stodola RK;  
 XX  
 DR MPI; 1998-159452/14.  
 DR P-PSDB; AAV85867.  
 XX  
 PI Streptococcus pneumoniae proteins and related DNA - useful for  
 PT screening compounds for antibacterial activity  
 XX  
 PS Claim 4; Page 101-102; 640pp; English.  
 XX  
 CC This invention describes novel isolated Streptococcus pneumoniae  
 CC polynucleotides (see AA296173-296494) and their encoded proteins (see  
 CC AA85792-186182). The DNA, vectors and host cells described in the  
 CC method of the invention are useful for the recombinant expression of the  
 CC polypeptides. The polypeptides are useful for treatment or prevention of  
 CC disease, or diagnosis of disease related to expression or activity of  
 CC such a polypeptide. They can also be used to screen for compounds which  
 CC interact with and inhibit or activate such a polypeptide. The  
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful  
 CC for inducing an immunological response in a mammal. The antagonists are  
 CC useful to inhibit such bacterial polypeptides. The polypeptides are  
 CC particularly useful to identify antimicrobial compounds and antibiotics.  
 CC They are also useful to determine their role in pathogenesis of  
 CC infection, dysfunction and disease.  
 XX  
 SQ Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 1634;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cgtcaggtcagcagt 21  
 |||||||||  
 DB 954 cgtcaggtcagcagt 967

RESULT 29  
 AAV37360  
 ID AAV37360 standard; DNA; 1634 BP.  
 XX  
 AC AAV37360;  
 XX  
 DT 13-OCT-1998 (first entry)  
 XX  
 DE Streptococcus pneumoniae coding region.  
 XX  
 KW coding region; ORF; open reading frame; antibacterial;  
 XX infection; prevention; meningitis; ss.  
 OS Streptococcus pneumoniae.  
 XX  
 PN MO9819689-A1.  
 PD 14-MAY-1998.  
 XX  
 PF 27-OCT-1997; 97WO-US19226.  
 XX  
 PR 01-NOV-1996; 96US-0029930.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
 XX Reid RH, Zarfos FN;  
 DR MPI; 1998-286586/25.  
 DR P-PSDB; AAW60963.

```

XX New isolated nucleic acids from Streptococcus pneumoniae - useful,
PT e.g. for identifying anti-bacterial(s) for treatment and prevention
PT of meningitis
PS Claim 1: Page 60; 130pp; English.
XX
CC The sequence is that of a coding region isolated from
CC S. pneumoniae. Its encoded protein, or agonists of it,
CC may be useful as an antibacterial for treatment or
CC (prevention of infection, specifically caused by S. pneumoniae
CC (particularly meningitis) but possibly also Helicobacter
CC pylori (ulcers and gastric cancer). It may be of particular
CC use before insertion of an in-dwelling device or any other
CC invasive procedure. The protein, or nucleic acid encoding
CC it, can also be used in vaccines to induce a cellular
CC and/or humoral immune response, or to screen for other
CC antibacterials. The DNA may also contain flanking sequences
CC that are potential sources of control elements for bacterial
CC gene expression. Detecting a sequence encoding the protein
CC can be used diagnostically, e.g. to detect a mutation for
CC serotyping or classifying infectious agents.
XX
SQ Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;
XX
Query Match          56.0%; Score 14; DB 19; Length 1634;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 cgtcaggtgcaggt 21
   |||||||
DB 954 cgtcaggtgcaggt 967
XX
RESULT 30
AAV44597/C
ID AAV44597 standard; DNA: 1758 BP.
XX
AC AAV44597;
XX
DT 24-NOV-1998 (first entry)
XX
DE Mouse uncoupling protein-2 UCP2 gene promoter region.
XX
KM Uncoupling protein-2; UCP2 gene; mouse; respiration;
KM thermogenesis; obesity; hyperinsulinaemia; glucose intolerance;
KM diabetes; syndrome X; hypothermia; wasting; cachexia; anorexia;
KM inflammation; fever; hyperthermia; gene therapy; diagnosis; ds.
XX
OS Mus sp.
XX
PN WO9831396-A1.
XX
PD 23-JUL-1998.
XX
PE 22-APR-1997; 97WO-US06864.
XX
PR 15-JAN-1997; 97US-0034960.
XX
PA (NARE-) CENT NAT RECH SCI CENT RECH SUR ENDOCRINOL.
PA (REGC) UNIV CALIFORNIA.
PA (UYDU-) UNIV DUKE.
XX
PI Bouillaud F, Collins SA, Riquier D, Seldin MF;
PI Surwit RS, Warden CH;
XX
DR WPI; 1998-413823/35.
XX
PT Method for treating disease associated with altered UCP-2 expression
PT - by administering agent which enhances or inhibits UCP-2 activity,
PT effectively to treat obesity, diabetes, fever, hyperthermia,
PT cachexia etc..

```

```

XX Example II; Fig 8a; 98pp; English.
PS
XX This is the nucleotide sequence of a fragment of the mouse
XX uncoupling protein-2 (UCP2) gene, corresponding to the promoter
XX region, obtained from genomic clone MMU2-I2. A DNA fragment from
XX -334 to +8600 bp of the UCP2 gene has also been sequenced (see
XX AAV44596). MMU2-I2 (deposited as T-1868) was cloned from a mouse
XX genomic library by screening with mouse UCP2 cDNA; the cDNA had
XX been obtained from a mouse muscle cDNA library using rat UCP1 cDNA
XX as probe. Human UCP2 cDNA and genomic sequences (see AAV44595,
XX AA44598-602 and AAV44629-38) are also provided, as well as a deduced
XX amino acid sequence (see AAV69166) for human UCP2. The invention
XX relates to a method for treating disease associated with altered
XX UCP2 expression. An agent which enhances UCP2 expression can be
XX used to treat obesity, diabetes, syndrome X, hypothermia,
XX hyperinsulinaemia, or glucose intolerance. An inhibitor of UCP2
XX is used to treat wasting, anorexia, inflammation, cachexia, fever
XX or hyperthermia. The invention also relates to diagnostic and drug
XX screening methodologies.
XX
SQ Sequence 1758 BP; 421 A; 401 C; 438 G; 487 T; 11 other;
XX
Query Match          56.0%; Score 14; DB 19; Length 1758;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 atctcgtcaggtg 16
   |||||||
DB 1072 ATCTCTGTCAGGTG 1059
XX
RESULT 31
AAZ36323/C
ID AAZ36323 standard; cDNA: 3720 BP.
XX
AC AAZ36323;
XX
DT 22-FEB-2000 (first entry)
XX
DE Mechanical stress induced cDNA encoding protein 405.
XX
KM Mechanical stress; gene therapy; protein 405; osteoporosis; bone density;
KM bone development; ss.
XX
OS Rattus sp.
XX
OS
XX
FH Key Location/Qualifiers
FH CDS 46..3177
FH /tag=a
FH /product="protein 405"
XX
PN WO9960164-A1.
XX
PD 25-NOV-1999.
XX
PE 14-MAY-1999; 99WO-US11066.
XX
PR 15-MAY-1998; 98US-0085673.
XX
PA (QUAR-) QUARK BIOTECH INC.
XX
PI Einat P, Mor O, Skallter R, Feinstein E, Faerman A;
PI WPI; 2000-053304/04.
XX
DR P-PSDB; AAV53669.
XX
PT Identification of stress induced genes for determining risk and
PT preventing, treating or controlling osteoporosis -
XX Claim 25; Fig 9A-M; 308pp; English.
XX

```

CC The present sequence encodes protein 405, which was identified  
 CC using the method of the invention after subjecting rat osteoblasts to  
 CC mechanical stress. Expression of the 405 gene was found to be  
 CC downregulated in response to mechanical stimulation and stress.  
 CC The specification describes a method for the identification of genes  
 CC responsive to a specific mechanical stress. The method comprises applying  
 CC the mechanical stress to an organism (tissue or cells extracting bone  
 CC cells), isolating the specific cellular fractions and extracting mRNA  
 CC from them, and differentially analysing the mRNA in comparison with  
 CC control samples. The method is used to identify genes whose expression  
 CC is responsive to a specific stress. The identified genes are employed in  
 CC determining risk associated with a physiological or disease state. The  
 CC risk determination methods are used for testing a medication for gene  
 CC therapy. These medications, or genes identified by the method of the  
 CC invention, are used for treating, preventing or controlling a bone  
 CC physiological or disease state (especially osteoporosis or bone density  
 CC or other factors causing or contributing to osteoporosis or its symptoms  
 CC or other conditions involved in mechanical stress or its lack. The  
 CC methods can also be used for advancing research or studies in bone  
 CC development.

XX  
 XX Sequence 3720 BP; 946 A; 991 C; 955 G; 828 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 3720;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gacgcctgcagtc 15  
 |||||  
 Db 574 GATCCTGTCAGGT 561

## RESULT 32

AAAI1992  
 ID AAAI1992 standard; DNA; 37856 BP.

XX  
 AC AAAI1992;

XX  
 DT 07-AUG-2000 (first entry)

XX  
 DE s. cellulorum DNA encoding polyketide and heteropolyketide enzymes.

XX  
 KM polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;

KW epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;

KW plant-protection; ds.

XX  
 OS Sorangium cellulosum.

XX  
 FH Key Location/Qualifiers  
 FT CDS complement (3398..6100)

FT  
 FT /tag= a  
 FT /product= "ORF1-tRNA synthetase"

FT  
 FT /note= "gtg start codon"

FT  
 FT complement (6374..7111)

FT  
 FT /tag= b  
 FT /product= "ORF2-monoxygenase"

FT  
 FT complement (8433..9550)

FT  
 FT /tag= c  
 FT /product= "ORF3-aminotransferase"

FT  
 FT /note= "AGT start codon given in the specification"

FT  
 FT 9855..11393

FT  
 FT /tag= d  
 FT /product= "ORF4-tyrosine/DOPA-decarboxylase"

FT  
 FT /note= "GTG start codon"

FT  
 FT 12212..13658

FT  
 FT /tag= e  
 FT /product= "ORF5-3-oxoacyl-ACP-reductase"

FT  
 FT /note= "ACC start codon"

FT  
 FT 15374..19984

FT  
 FT /tag= f  
 FT /product= "ORF6-polyketide synthase"

FT  
 FT 20003..27889

FT  
 FT CDS

FT  
 FT /tag= g  
 FT /product= "ORF7-peptide synthetase"

FT  
 FT 28251..29400

FT  
 FT /tag= h  
 FT /product= "ORF8-transpeptidase"

FT  
 FT complement (30040..31720)

FT  
 FT /tag= i  
 FT /product= "ORF9-regulation element"

FT  
 FT /note= "CGC stop codon"

FT  
 FT 31982..32932

FT  
 FT /tag= j  
 FT /product= "ORF10-transcription regulator"

FT  
 FT 33128..33613

FT  
 FT /tag= k  
 FT /product= "ORF11-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 33661..34077

FT  
 FT /tag= l  
 FT /product= "ORF12-regulation element"

FT  
 FT complement (35255..35616)

FT  
 FT /tag= m  
 FT /product= "ORF13-transcription regulator"

FT  
 FT complement (35730..36242)

FT  
 FT /tag= n  
 FT /product= "ORF14-transcription regulator"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= o  
 FT /product= "ORF15-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= p  
 FT /product= "ORF16-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= q  
 FT /product= "ORF17-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= r  
 FT /product= "ORF18-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= s  
 FT /product= "ORF19-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= t  
 FT /product= "ORF20-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= u  
 FT /product= "ORF21-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= v  
 FT /product= "ORF22-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= w  
 FT /product= "ORF23-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= x  
 FT /product= "ORF24-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= y  
 FT /product= "ORF25-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= z  
 FT /product= "ORF26-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= aa  
 FT /product= "ORF27-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ab  
 FT /product= "ORF28-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ac  
 FT /product= "ORF29-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ad  
 FT /product= "ORF30-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ae  
 FT /product= "ORF31-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= af  
 FT /product= "ORF32-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ag  
 FT /product= "ORF33-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ah  
 FT /product= "ORF34-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ai  
 FT /product= "ORF35-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= aj  
 FT /product= "ORF36-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ak  
 FT /product= "ORF37-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= al  
 FT /product= "ORF38-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= am  
 FT /product= "ORF39-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= an  
 FT /product= "ORF40-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ao  
 FT /product= "ORF41-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ap  
 FT /product= "ORF42-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= aq  
 FT /product= "ORF43-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ar  
 FT /product= "ORF44-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= as  
 FT /product= "ORF45-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= at  
 FT /product= "ORF46-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= au  
 FT /product= "ORF47-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= av  
 FT /product= "ORF48-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= aw  
 FT /product= "ORF49-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ax  
 FT /product= "ORF50-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ay  
 FT /product= "ORF51-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= az  
 FT /product= "ORF52-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= ba  
 FT /product= "ORF53-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bb  
 FT /product= "ORF54-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bc  
 FT /product= "ORF55-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bd  
 FT /product= "ORF56-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= be  
 FT /product= "ORF57-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bf  
 FT /product= "ORF58-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bg  
 FT /product= "ORF59-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bh  
 FT /product= "ORF60-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bi  
 FT /product= "ORF61-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bj  
 FT /product= "ORF62-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bk  
 FT /product= "ORF63-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bl  
 FT /product= "ORF64-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bm  
 FT /product= "ORF65-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bn  
 FT /product= "ORF66-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bo  
 FT /product= "ORF67-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bp  
 FT /product= "ORF68-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bq  
 FT /product= "ORF69-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= br  
 FT /product= "ORF70-regulation element"

FT  
 FT /note= "GTG start codon"

FT  
 FT 31982..32932

FT  
 FT /tag= bs

```

XX AC AAC81914;
XX AC
XX DT 27-FEB-2001 (first entry)
XX DE Chlamydia pneumoniae genome DNA.
XX KM Genome: diagnosis; vaccine: ds.
XX OS Chlamydia pneumoniae.
XX PN W0200027994-A2.
XX PD 18-MAY-2000.
XX PE 12-NOV-1999; 99WO-US26923.
XX PR 12-NOV-1998; 98US-0108279.
XX PR 08-APR-1999; 99US-0128606.
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Stephens R, Mitchell W, Kalman S, Davis R;
XX DR WPI; 2000-376516/32.
XX PT Isolated nucleic acid for use in diagnostic and analytical methods
XX PT encodes genomic sequence of Chlamydia pneumoniae .
XX PS Claim 2; Page 128-320; 320pp; English.
XX CC This invention describes a novel nucleic acid (N1) encoding a Chlamydia
XX CC pneumoniae protein (PI), given in the specification. The isolated nucleic
XX CC acid is useful for diagnostic and analytical methods, such as,
XX CC hybridization-based assays or amplification-based assays. The protein may
XX CC be used for diagnostic purposes, for their enzymatic or structural
XX CC activity, or as a vaccine. The invention also describes (1) a probe
XX CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
XX CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
XX CC cassette comprising N1 under the transcriptional regulation of a
XX CC transcriptional initiation region functional in an expression host, and a
XX CC cassette of (3) as part of an extrachromosomal element or integrated into
XX CC the genome of a host cell as a result of induction of the expression
XX CC cassette into the host cell, and the cellular progeny of the host cell;
XX CC (5) a method for producing a PI comprising growing a cell of (4) where
XX CC the protein is expressed and isolating the protein free of other
XX CC proteins; (6) a purified polypeptide composition comprising at least 50
XX CC weight % of PI; and (7) a monoclonal antibody binding specifically to the
XX CC peptide of (6).
XX CC
XX CC Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;
XX CC
XX CC Query Match 56.0%; Score 14; DB 21; Length 273254;
XX CC Best Local Similarity 100.0%; Pred. No. 44;
XX CC Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX CC
XX CC 2 gatccctgcaggt 15
XX CC |||||||||||||
XX CC Db 32246 gatccctgcaggt 32259
XX CC
XX CC RESULT 34
XX CC AAV30458
XX CC ID AAV30458 standard; DNA; 534720 BP.
XX CC
XX CC AAV30458;
XX CC
XX CC 14-OCT-1998 (first entry)
XX CC
XX CC Rhizobium species plasmid pNGR234a.
XX CC

```

```

XX OS Rhizobium sp.
XX KM Symbolosis; open reading frame; ORF; plasmid; vector; transportation;
XX KM degradation; metabolism; host range; nitrogen fixation; nodulation;
XX KM legume; plant; ds.
XX KM
XX KM Key
XX KM CDS
XX KM Location/Qualifiers
XX KM 417796..418671
XX KM /tag= a
XX KM /standard_name= "ORF K1"
XX KM /product= "oligopeptide permease"
XX KM /note= "homologous to the OppC gene"
XX KM 418673..419680
XX KM /tag= b
XX KM /standard_name= "ORF K2"
XX KM /product= "oligopeptide permease"
XX KM /note= "homologous to the OppD gene"
XX KM 419677..420738
XX KM /tag= c
XX KM /standard_name= "ORF K3"
XX KM /product= "oligopeptide permease"
XX KM /note= "homologous to the OppF gene"
XX KM 420774..422159
XX KM /tag= d
XX KM /standard_name= "ORF K4"
XX KM /product= "encapsulation-like protein"
XX KM /note= "homologous to the CapA gene"
XX KM 422628..424031
XX KM /tag= e
XX KM /standard_name= "ORF K5"
XX KM /product= "aminotransferase-like protein"
XX KM /note= "homologous to the BioA gene"
XX KM 424056..425594
XX KM /tag= f
XX KM /standard_name= "ORF K6"
XX KM /product= "(semi)aldehyde dehydrogenase-like protein"
XX KM /note= "complement (426949..428028)"
XX KM /tag= g
XX KM /standard_name= "ORF K7"
XX KM /product= "transposase homologue"
XX KM /note= "homologous to the Tnp gene"
XX KM 428292..429623
XX KM /tag= h
XX KM /standard_name= "ORF K8"
XX KM /product= "glutamate dehydrogenase-like protein"
XX KM /note= "homologous to the GluD1 gene"
XX KM /note= "complement (430538..431284)"
XX KM /tag= i
XX KM /standard_name= "ORF K9"
XX KM /product= "transposase homologue"
XX KM /note= "complement (431296..432840)"
XX KM /tag= j
XX KM /standard_name= "ORF K10"
XX KM /product= "transposase homologue"
XX KM /note= "homologous to the Tnp gene"
XX KM /note= "complement (433880..434110)"
XX KM /tag= k
XX KM /standard_name= "ORF K11"
XX KM /product= "protein of unknown function"
XX KM /note= "homologous to the FixV gene"
XX KM /note= "complement (434107..434433)"
XX KM /tag= l
XX KM /standard_name= "ORF K12"
XX KM /product= "protein of unknown function"
XX KM /note= "complement (434517..434711)"
XX KM /tag= m
XX KM /standard_name= "ORF K13"
XX KM /product= "ferrodoxin/ferrodoxin-like protein"
XX KM /note= "homologous to the FdxN gene"
XX KM /note= "complement (434753..436234)"
XX KM /tag= n
XX KM /standard_name= "ORF K14"
XX KM /gene= "nifB"
XX KM

```

```

FT      /product= "protein involved in Femo co-factor
FT      biosynthesis"
FT      complement (436460..438130)
FT      /tag= o
FT      /standard_name= "ORF K15"
FT      /gene= "nifA"
FT      /product= "positive regulator of nif, fix and other
FT      genes"
FT      complement (438297..438590)
FT      /tag= p
FT      /standard_name= "ORF K16"
FT      /gene= "fixX"
FT      /product= "protein required for nitrogenase activity"
FT      complement (438605..439912)
FT      /tag= q
FT      /standard_name= "ORF K17"
FT      /gene= "fixC"
FT      /product= "protein required for nitrogenase activity"
FT      complement (439923..441032)
FT      /tag= i
FT      /standard_name= "ORF K18"
FT      /gene= "fixB"
FT      /product= "protein required for nitrogenase activity"
FT      complement (441042..441899)
FT      /tag= s
FT      /standard_name= "ORF K19"
FT      /gene= "fixA"
FT      /product= "protein required for nitrogenase activity"
FT      complement (442316..442636)
FT      /tag= t
FT      /standard_name= "ORF K20"
FT      /product= "protein of unknown function"
FT      complement (44313..443879)
FT      /tag= u
FT      /standard_name= "ORF K21"
FT      /product= "protein of unknown function"
FT      complement (44337..445029)
FT      /tag= v
FT      /standard_name= "ORF K22"
FT      /product= "Ferredoxin-like protein"
FT      /note= "homologous to the NifQ gene"
FT      complement (445088..446602)
FT      /tag= w
FT      /standard_name= "ORF K23"
FT      /gene= "dctA"
FT      /product= "C4-dicarboxylate transport protein"
FT      /note= "homologous to the DctAI gene"
FT      complement (446599..447843)
FT      /tag= x
FT      /standard_name= "ORF L1"
FT      /product= "cytochrome P450-like protein"
FT      /note= "homologous to the CamC gene"
FT      complement (447844..448500)
FT      /tag= y
FT      /standard_name= "ORF L2"
FT      /product= "gamma-hexachlorocyclohexane-dechlorinase-like
FT      protein"
FT      /note= "homologous to the LinA gene"
FT      complement (448497..450203)
FT      /tag= z
FT      /standard_name= "ORF L3"
FT      /product= "putative protein with degradative function"
FT      complement (450341..451396)
FT      /tag= aa
FT      /standard_name= "ORF L4"
FT      /product= "luciferase alpha-subunit-like protein"
FT      /note= "homologous to the LuxA gene"
FT      complement (452980..454454)
FT      /tag= ab
FT      /standard_name= "ORF L6"
FT      /gene= "nifD"
FT      /product= "alpha-subunit of fmo protein of nitrogenase"
FT      complement (454590..456131)

```

```

FT      /tag= ac
FT      /standard_name= "ORF L7"
FT      /gene= "nifK"
FT      /product= "beta-subunit of fmo protein of nitrogenase"
FT      complement (456187..457677)
FT      /tag= ad
FT      /standard_name= "ORF L8"
FT      /product= "protein involved in fmo co-factor
FT      biosynthesis"
FT      /note= "homologous to the NifB gene"
FT      complement (457687..459096)
FT      /tag= ae
FT      /standard_name= "ORF L9"
FT      /product= "protein involved in fmo co-factor
FT      biosynthesis"
FT      /note= "homologous to the fixF gene"
FT      complement (459093..459575)
FT      /tag= af
FT      /standard_name= "ORF L10"
FT      /product= "protein of unknown function"
FT      /note= "homologous to the Nifx gene"
FT      complement (459579..460067)
FT      /tag= ag
FT      /standard_name= "ORF L11"
FT      /product= "protein of unknown function"
FT      complement (460501..460920)
FT      /tag= ah
FT      /standard_name= "ORF L12"
FT      /product= "protein similar to part of the Fe protein
FT      of nitrogenase"
FT      /note= "homologous to the NifH gene"
FT      complement (461228..461545)
FT      /tag= ai
FT      /standard_name= "ORF L13"
FT      /product= "protein of unknown function"
FT      complement (461301..464739)
FT      /tag= aj
FT      /standard_name= "ORF L14"
FT      /product= "peptidase-like protein"
FT      /note= "homologous to the Df-MPP gene"
FT      complement (464736..466079)
FT      /tag= ak
FT      /standard_name= "ORF L15"
FT      /product= "processing protease-like protein"
FT      /note= "homologous to the PP gene"
FT      complement (466590..467021)
FT      CDS

```

Query Match 56.0%; Score 14; DB 19; Length 534720;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      10 tcaggtgcaggtca 23
DB 435493 tcaggtgcaggtca 435506

```

RESULT 35  
AAV30459  
ID AAV30459 standard; DNA; 536165 BP.  
AC AAV30459;  
XX  
DT 06-JUL-1999 (first entry)  
XX  
DE Rhizobium species symbiotic plasmid pNGR234.  
XX  
KW Sympliosis; open reading frame; ORF; plasmid; vector; transportation;  
KW degradation; metabolism; host range; nitrogen fixation; nodulation;  
KW legume; plant; ds.  
XX  
XX Rhizobium sp.  
OS  
XX  
PN W09802560-A2.

```

XX 22-JAN-1998.
PD 10-JUL-1997; 97WO-IB00950.
XX 20-MAY-1997; 97GB-0010395.
XX 12-JUL-1996; 96EP-0730001.
XX (MOLE-) INST MOLECULAR BIOTECHNOLOGY.
PA (BIOL-) LAB BIOLOGIE MOLECULAIRE PLANTES SUPERIE.
XX Broughton WJ, Freilberg CB, Perret XP, Rosenthal A;
XX WPI; 1998-110606/10.
DR
XX
XX New isolated symbiotic plasmid from Rhizobium sp. NGR234 - used to
PT develop products for modifying plant characteristics, e.g. nitrogen
PT fixation, synthesis of compounds and stress response
XX
XX Claim 1; Fig 3; 228pp; English.
PS
XX
XX This is the nucleotide sequence of the plasmid pNGR234a isolated from
CC Rhizobium sp. NGR234. Open reading frames (ORF) derivable from the
CC nucleotide sequence are claimed. The nucleotide sequences or ORFs can
CC be used e.g. in the transportation of compounds to and from an organism
CC which is a host to at least one of the nucleotide sequences, ORFs or
CC proteins, the degradation and/or metabolism of organic, inorganic,
CC natural or xenobiotic substances in a host organism or the modification
CC of the host range, nitrogen fixation abilities; for obtaining a synthetic
CC minimal set of ORFs required for functional Rhizobium-legume symbiosis,
CC especially for nodulation efficiency on host plants.
XX
XX Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 other;
SQ
Query Match 56.0%; Score 14; DB 19; Length 536165;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 tcaggtcagtcag 23
Db 435493 tcaggtcagtcag 435506
|||||
RESULT 36
AA219643
ID AA219643 standard; RNA; 66 BP.
XX
XX AC AA219643;
XX
XX 08-NOV-1999 (first entry)
XX
XX DE Complement system protein C1q RNA binding ligand 29.
XX
XX KW Complement system protein; C1q; C5; C3; RNA ligand; neutrotropic; stroke;
XX neuroprotective; anti-asthmatic; cerebroprotective; immunosuppressive;
XX diagnostic; prevention; treatment; complement protein-related disease;
XX Alzheimer's disease; renal disease; transplant rejection; asthma; ss.
XX
XX OS Synthetic.
XX
XX PN WO9941271-A1.
XX
XX PD 19-AUG-1999.
XX
XX PE 05-FEB-1999; 99WO-US02597.
XX
XX PR 29-SEP-1998; 98US-0163025.
XX 12-FEB-1998; 98US-0023228.
XX
XX PA (NEXS-) NEXSTAR PHARM INC.
XX
XX PI Biesecker G, Gold L;

```

```

XX WPI; 1999-527357/44.
DR
XX New Nucleic Acid ligand to complement protein C5, useful in the
PT treatment of Alzheimer's disease, renal diseases, transplant
PT rejection, stroke and asthma
XX
XX Example 6; Page 45; 120pp; English.
XX
XX This invention describes novel purified and isolated non-naturally
CC occurring RNA ligands (I) to complement system proteins C5, C1q and C3.
CC The products of the invention have neutrotropic, neuroprotective,
CC anti-asthmatic, cerebroprotective and immunosuppressive activity. The
CC new ligands are useful as diagnostic agents, and pharmaceutical agents
CC for prevention and treatment of complement protein-related diseases,
CC including Alzheimer's disease, renal diseases, transplant rejection,
CC stroke and asthma.
XX
XX Sequence 66 BP; 11 A; 19 C; 23 G; 13 U; 0 other;
SQ
Query Match 52.0%; Score 13; DB 20; Length 66;
Best Local Similarity 76.9%; Pred. No. 2.8e+02;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgatcctcgctcag 13
Db 41 cgatcctcgctcag 53
|||||
RESULT 37
AA85925/C
ID AA85925 standard; DNA; 123 BP.
XX
XX AC AA85925;
XX
XX 13-SEP-1999 (first entry)
XX
XX DE Oligonucleotide used to produce heavy chain variable region of Ig NEW.
XX
XX KW Light chain variable region; interleukin-4; IL-4; antibody 389;
XX chimeric antibody; humanized antibody; IL-4 mediated allergic reaction;
XX immunoglobulin E-mediated allergic reaction; allergic rhinitis;
XX conjunctivitis; atopic dermatitis; atopic asthma; anaphylactic shock;
XX rheumatoid arthritis; host-versus-graft disease; renal disease;
XX allergy; Ig NEW; ss.
XX
XX OS Synthetic.
XX
XX PN US5928904-A.
XX
XX PD 27-JUL-1999.
XX
XX PE 07-JUN-1995; 95US-0483632.
XX
XX PR 07-JUN-1995; 95US-0483632.
XX 07-SEP-1993; 93US-0117366.
XX 14-OCT-1993; 93US-0136783.
XX 07-SEP-1994; 94WO-US10308.
XX
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX (SMIK ) SMITHKLINE BEECHAM PLC.
XX
XX PI Gross MS, Holmes SD, Sylvester DR;
XX
XX DR WPI; 1999-429500/36.
XX
XX PT New DNA molecules encoding recombinant antibodies useful for
XX treating IL4-mediated conditions
XX
XX Disclosure; Columns 61-63; 50pp; English.
XX
XX CC The specification describes chimeric and humanised IL-4

```



monoclonal antibodies. The antibodies of the invention are used in therapeutic and pharmaceutical compositions for treating IL-4 mediated and Immunoglobulin E-mediated allergic reactions e.g. allergic rhinitis, conjunctivitis, atopic dermatitis, anaphylactic shock, rheumatoid arthritis, host-versus-graft disease and renal disease. They are also useful in the diagnosis of an allergy or condition associated with excess IL-4 production through the measurement e.g. by ELISA of circulating endogenous IL-4 levels in humans. Oligonucleotides AAX85925-28 were used to produce the heavy chain variable region of Ig NEW, in the course of the invention.

Sequence 123 BP; 25 A; 37 C; 41 G; 20 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 123;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcagtcag 24  
|||||  
DB 70 AGGTGCAGTCAG 58

RESULT 38  
AAC00049/c  
ID AAC00049 standard; CDNA: 276 BP.

AAC00049;  
06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 47.

Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST ) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

P-PSDB; AAC00043.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

Claim 1; SEQ ID 47; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. An ORF has been identified within the sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

XX  
Sequence 276 BP; 55 A; 93 C; 69 G; 59 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 276;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtgcagtcagc 25  
|||||  
DB 143 GGTGCAGTCAGC 131

RESULT 39  
AAC08252/c  
ID AAC08252 standard; CDNA: 351 BP.

AAC08252;

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 12327.

Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST ) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

Claim 1; SEQ ID 12327; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. No ORF has yet been conclusively identified within the present sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 351 BP; 58 A; 105 C; 122 G; 66 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 351;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctctgcagtg 16  
|||||  
DB 220 TCCTCGTCAGTG 208

RESULT 40  
AAK85929/C  
ID AAK85929 standard; DNA: 363 BP.  
XX  
AC AAK85929;  
XX  
DE 13-SEP-1999 (first entry)  
XX  
DE DNA encoding the heavy chain variable region of Ig NW.  
XX  
KM Heavy chain variable region; interleukin-4; IL-4; antibody 3B9;  
KM chimeric antibody; humanised antibody; IL-4 mediated allergic reaction;  
KM immunoglobulin E-mediated allergic reaction; allergic rhinitis;  
KM conjunctivitis; atopic dermatitis; atopic asthma; anaphylactic shock;  
KM rheumatoid arthritis; host-versus-graft disease; renal disease;  
KM allergy; Ig NEW; ss.  
XX  
OS Synthetic.  
XX  
PN US5928904-A.  
XX  
PD 27-JUL-1999.  
XX  
PF 07-JUN-1995; 9505-0483632.  
XX  
PR 07-JUN-1995; 9505-0483632.  
PR 07-SEP-1993; 9305-0117366.  
PR 14-OCT-1993; 9305-0136783.  
PR 07-SEP-1994; 94WO-US10308.  
XX  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PA (SMK ) SMITHKLINE BEECHAM PLC.  
XX  
PI Gross MS, Holmes SD, Sylvester DR;  
PI WPI: 1999-429500/36.  
DR P-PSDB; AAY23780.  
XX  
PT New DNA molecules encoding recombinant antibodies useful for  
PT treating IL4-mediated conditions  
XX  
PS Disclosure; Columns 63-65; 50pp; English.  
XX  
CC The specification describes chimeric and humanised IL-4  
CC monoclonal antibodies. The antibodies of the invention are used in  
CC therapeutic and pharmaceutical compositions for treating IL-4 mediated  
CC and immunoglobulin E-mediated allergic reactions e.g. allergic rhinitis,  
CC conjunctivitis, atopic dermatitis, atopic asthma, anaphylactic shock,  
CC rheumatoid arthritis, host-versus-graft disease and renal disease.  
CC They are also useful in the diagnosis of an allergy or condition  
CC associated with excess IL-4 production through the measurement e.g. by  
CC ELISA of circulating endogenous IL-4 levels in humans. The present  
CC sequence encodes the heavy chain variable region of Ig NEW, and is  
CC used in the course of the invention.  
XX  
SQ Sequence 363 BP; 79 A; 116 C; 111 G; 57 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 363;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcagtcag 24  
DB 70 AGGTGCAGTCAG 58  
|||||

RESULT 41  
AAK79527/C  
ID AAK79527 standard; cDNA; 363 BP.  
XX

AC AAK79527;  
XX  
DT 11-AUG-1999 (first entry)  
XX  
DE Heavy chain coding sequence for humanised 3B9 antibody.  
XX  
KM Antibody; interleukin-4; IL4; immunoglobulin E; IgE mediated disease;  
KM allergic disorder; allergic rhinitis; conjunctivitis; atopic dermatitis;  
KM atopic asthma; anaphylactic shock; cell proliferation regulator; therapy;  
KM autoimmune disease; graft versus host disease; ss.  
XX  
OS Synthetic.  
XX  
PN US5914110-A.  
XX  
PD 22-JUN-1999.  
XX  
PF 07-JUN-1995; 9505-0483636.  
XX  
PR 07-JUN-1995; 9505-0483636.  
PR 07-SEP-1993; 9305-0117366.  
PR 14-OCT-1993; 9305-0136783.  
PR 07-SEP-1994; 94WO-US10308.  
XX  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PA (SMK ) SMITHKLINE BEECHAM PLC.  
XX  
PI Gross MS, Holmes SD, Sylvester DR;  
PI WPI: 1999-370482/31.  
DR P-PSDB; AAY18122.  
XX  
PT Recombinant IL4 antibodies  
PT Example 3; Column 63-64; 50pp; English.  
XX  
PS This sequence encodes the heavy chain of the humanised 3B9  
PS antibody of the invention. The antibody is a chimeric or  
CC humanised interleukin-4 (IL4) monoclonal antibody for the treatment of  
CC immunoglobulin E (IgE) mediated diseases. The antibodies are useful for  
CC the treatment of allergic disorders such as allergic rhinitis,  
CC conjunctivitis, atopic dermatitis, atopic asthma and anaphylactic shock.  
CC The antibodies are also useful for regulating B and T cell proliferation  
CC and as such are useful in the treatment of autoimmune diseases and graft  
CC versus host disease.  
XX  
SQ Sequence 363 BP; 79 A; 116 C; 111 G; 57 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 363;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcagtcag 24  
DB 70 AGGTGCAGTCAG 58  
|||||

RESULT 42  
AAF22004  
ID AAF22004 standard; DNA; 375 BP.  
XX  
AC AAF22004;  
XX  
DT 27-MAR-2001 (first entry)  
XX  
DE Human breast and ovarian cancer associated antigen gene SEQ ID 391.  
XX  
KM Human; breast cancer; ovarian cancer; cytosolic; immunosuppressive;  
KM neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
KM antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;  
KM antibacterial; antifungal; antiparasitic; cardiant; immune disorder;  
KM Addison's disease; allergy; autoimmune haemolytic anaemia;

KM autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
 KM multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
 KM cardiovascular disorder; wound healing; neurological disease; ds.  
 OS Homo sapiens.  
 XX MO200055173-A1.  
 PN 21-SEP-2000.  
 PD 08-MAR-2000; 2000WO-US05881.  
 PF 12-MAR-1999; 99US-0124270.  
 XX (HUMA-) HUMAN GENOME SCT INC.  
 PA Rosen CA, Ruben SM;  
 PI WPI: 2000-611515/58.  
 DR P-PSDB; AAB59101.  
 XX New human breast and ovarian cancer associated gene sequences and the  
 PT polypeptides encoded by these genes, useful in the prevention,  
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
 PT disorders and neurological diseases -  
 PS Claim 1; Page 823; 1299pp; English.  
 XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
 CC associated with breast and ovarian cancer. Included in the invention are  
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
 CC isolation and characterisation of the DNA and protein sequences of the  
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
 CC neurotropic; neuroproliferative; antiviral; anti-allergic; hepatotropic;  
 CC antibacterial; anti-inflammatory; antitumor; antileukemic; anticonvulsant;  
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The  
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
 CC particularly breast and ovarian cancer. The nucleic acid sequences,  
 CC proteins, agonists and antagonists may also be used in the diagnosis,  
 CC prevention and treatment of immune disorders e.g. Addison's disease,  
 CC allergies, autoimmune hemolytic anaemia, autoimmune thyroiditis,  
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
 CC arthritis and ulcerative colitis; cardiovascular disorders such as  
 CC myocardial ischaemias; wound healing; neurological diseases such as  
 CC cerebral anoxia and epilepsy; and infectious diseases.  
 XX Sequence 375 BP; 79 A; 77 C; 129 G; 71 T; 19 other:  
 SQ

Query Match 52.0%; Score 13; DB 21; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 gtcaagtgcaagt 21  
 |||||||||  
 DB 171 gtcaagtgcaagt 183

RESULT 43  
 AAC01754/c  
 ID AAC01754 standard; cDNA; 375 BP.  
 XX AAC01754;  
 AC  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 1752.  
 XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KM gene therapy; chromosome mapping; ss.  
 XX

OS Homo sapiens.  
 XX  
 PN EP103401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST) GENSET.  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 DR WPI: 2000-500381/45.  
 XX P-PSDB; AAG01748.  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PS Claim 1; SEQ ID 1752; 71pp + CD-ROM; English.  
 XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.  
 XX Sequence 375 BP; 66 A; 121 C; 117 G; 62 T; 9 other:  
 SQ

Query Match 52.0%; Score 13; DB 21; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tccctgcagtg 16  
 |||||||||  
 DB 182 TCCTGTCAGGTG 170

RESULT 44  
 AAT61237/c  
 ID AAT61237 standard; DNA; 378 BP.  
 XX AAT61237;  
 AC  
 XX  
 DT 13-MAY-1997 (first entry)  
 XX  
 DE Human anti-RSV monoclonal antibody RF-2 VH domain DNA.  
 XX Monoclonal antibody; MAb; RF-1; RF-2; respiratory syncytial virus;  
 KM RSV; fusion protein; F-protein; vaccine; immunotherapy; therapy;  
 KM Epstein Barr virus; immortalisation; recombinant antibody; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9640252-A1.  
 XX  
 PD 19-DEC-1996.  
 XX  
 DE 06-JUN-1996; 96WO-US10070.  
 PF 07-JUN-1995; 95US-0488376.  
 XX

```

XX PA (IDEC-) IDEC PHARM CORP.
XX PI Brame P, Chamat SS, Heard CJ, Newman RA, Pan L;
XX PI Walsh EE;
XX DR WPI: 1997-099892/09.
XX P-PSDB: AAW11625.
XX PT Human monoclonal antibody specific for respiratory syncytial virus
XX PT fusion protein - used for the prevention and treatment of RSV
XX PT infection
XX PS Claim 9; Fig 7b; 85pp; English.
XX CC Nucleic acids (AAT61236-39) respectively code for the RF-1 VL domain,
XX CC RF-2 VH domain, RF-2 VL domain and RF-1 VH domain (AAW11634-37) of
XX CC human monoclonal antibodies (hMabs) RF-1 and RF-2, which show high
XX CC affinity for respiratory syncytial virus (RSV) fusion protein. The
XX CC hMabs were obd. by: antigen priming of naive human splenocytes in
XX CC vitro; transferal of the primed splenocytes to a SCID mouse;
XX CC boosting with antigen; isolating human antibody-producing B-cells
XX CC from the SCID mice; and immortalisation of the B-cells using
XX CC Epstein Barr virus. DNA encoding the RF-1 and RF-2 heavy and
XX CC light variable domains can be incorporated into vectors (e.g.
XX CC NEOSPLA) and used to transfect eukaryotic (e.g. CHO) cells (see
XX CC also AAT61240-42, AAT61279) to provide a constant, stable supply of
XX CC anti-RSV F-protein hMabs for use in the treatment or prevention of
XX CC RSV infection.
XX CC (N.B. Fig 7b is referred to as the RF-1 VH domain in the
XX CC specification, but is identical to the RF-2 VH domain of Fig 11b).
XX SQ Sequence 378 BP; 91 A; 114 C; 90 G; 83 T; 0 other;

Query Match          52.0%; Score 13; DB 18; Length 378;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcaggtcag 24
   |||
DB 70 AGGTGCAGGTCTAG 58

RESULT 45
AAT61239/c
ID AAT61239 standard; DNA: 378 BP.
XX AAT61239;
AC AAT61239;
XX 13-MAY-1997 (first entry)
XX DE Human anti-RSV monoclonal antibody RF-1 VH domain DNA.
XX KM Monoclonal antibody; Mab; RF-1; RF-2; respiratory syncytial virus;
XX KM RSV; fusion protein; F-protein; vaccine; immunotherapy; therapy;
XX KM Epstein Barr virus; immortalisation; recombinant antibody; ss.
XX OS Homo sapiens.
XX PN WO9640252-A1.
XX PD 19-DEC-1996.
XX PF 06-JUN-1996; 96WO-US10070.
XX PR 07-JUN-1995; 95US-0488376.
XX PA (IDEC-) IDEC PHARM CORP.
XX PI Brame P, Chamat SS, Heard CJ, Newman RA, Pan L;
XX PI Walsh EE;

```

```

DR WPI: 1997-099892/09.
XX PT Human monoclonal antibody specific for respiratory syncytial virus
XX PT fusion protein - used for the prevention and treatment of RSV
XX PT infection
XX PS Claim 9; Fig 8b; 85pp; English.
XX CC Nucleic acids (AAT61236-39) respectively code for the RF-1 VL domain,
XX CC RF-2 VH domain, RF-2 VL domain and RF-1 VH domain (AAW11634-37) of
XX CC human monoclonal antibodies (hMabs) RF-1 and RF-2, which show high
XX CC affinity for respiratory syncytial virus (RSV) fusion protein. The
XX CC hMabs were obd. by: antigen priming of naive human splenocytes in
XX CC vitro; transferal of the primed splenocytes to a SCID mouse;
XX CC boosting with antigen; isolating human antibody-producing B-cells
XX CC from the SCID mice; and immortalisation of the B-cells using
XX CC Epstein Barr virus. DNA encoding the RF-1 and RF-2 heavy and
XX CC light variable domains can be incorporated into vectors (e.g.
XX CC NEOSPLA) and used to transfect eukaryotic (e.g. CHO) cells (see
XX CC also AAT61240-42, AAT61279) to provide a constant, stable supply of
XX CC anti-RSV F-protein hMabs for use in the treatment or prevention of
XX CC RSV infection.
XX CC (N.B. Fig 8b is referred to as the RF-2 VH domain in the
XX CC specification, but is identical to the RF-1 VH domain of Fig 9b).
XX SQ Sequence 378 BP; 86 A; 110 C; 97 G; 85 T; 0 other;

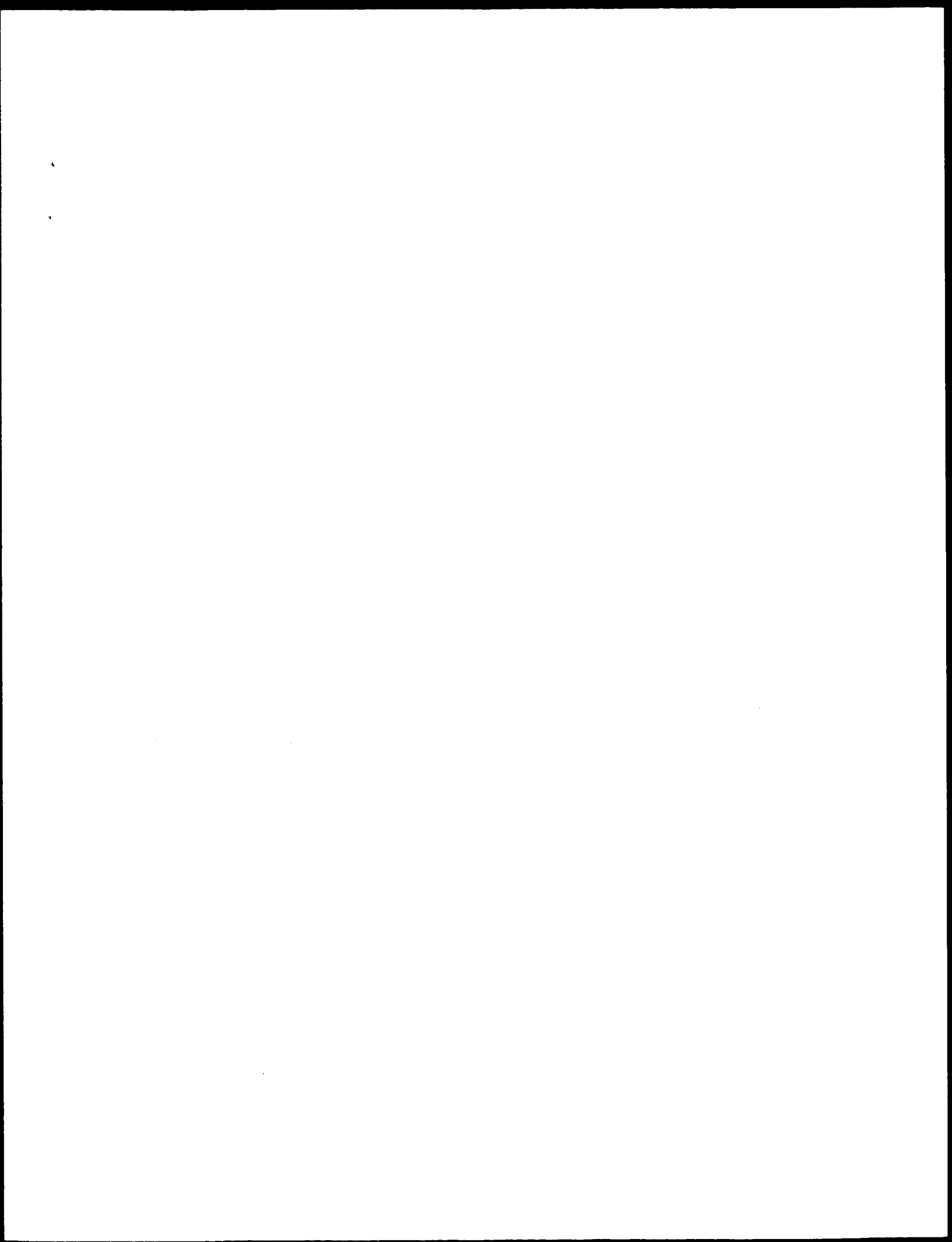
Query Match          52.0%; Score 13; DB 18; Length 378;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtgcaggtcagc 25
   |||
DB 69 GGTGCAGGTCTAGC 57

```

Search completed: October 9, 2001, 15:53:33  
 Job time: 15174 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:41 ; Search time 218.82 Seconds  
(without alignments)  
21.629 Million cell updates/sec

Title: US-09-396-196f-9

Perfect score: 25

Sequence: 1 cgatcctcgtaagtcagtcagtcagc 25

Scoring table: OLIGO\_MNC

Searched: 324599 seqs, 9465562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 11931

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A-COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B-COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A-COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B-COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS-COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result | Score | Query | Match | Length | DB                | ID | Description       |
|--------|-------|-------|-------|--------|-------------------|----|-------------------|
| No.    |       |       |       |        |                   |    |                   |
| 1      | 25    | 100.0 | 1041  | 2      | US-08-401-068-7   |    | Sequence 7, Appl  |
| 2      | 25    | 100.0 | 1041  | 2      | US-08-846-338-7   |    | Sequence 7, Appl  |
| 3      | 25    | 100.0 | 5872  | 3      | US-08-411-768B-1  |    | Sequence 1, Appl  |
| 4      | 25    | 100.0 | 5872  | 3      | US-08-411-768B-6  |    | Sequence 1, Appl  |
| 5      | 14    | 56.0  | 613   | 3      | US-08-545-809A-5  |    | Sequence 5, Appl  |
| 6      | 14    | 56.0  | 630   | 3      | US-08-545-809A-10 |    | Sequence 10, Appl |
| 7      | 13    | 52.0  | 66    | 3      | US-09-023-228B-96 |    | Sequence 96, Appl |
| 8      | 13    | 52.0  | 123   | 2      | US-08-483-636-59  |    | Sequence 59, Appl |
| 9      | 13    | 52.0  | 123   | 2      | US-08-483-632-59  |    | Sequence 59, Appl |
| 10     | 13    | 52.0  | 363   | 2      | US-08-483-636-63  |    | Sequence 63, Appl |
| 11     | 13    | 52.0  | 363   | 2      | US-08-483-632-63  |    | Sequence 63, Appl |
| 12     | 13    | 52.0  | 378   | 1      | US-08-488-376-13  |    | Sequence 13, Appl |
| 13     | 13    | 52.0  | 378   | 1      | US-08-488-376-15  |    | Sequence 15, Appl |
| 14     | 13    | 52.0  | 378   | 2      | US-08-634-223-13  |    | Sequence 13, Appl |
| 15     | 13    | 52.0  | 378   | 2      | US-08-634-223-15  |    | Sequence 15, Appl |
| 16     | 13    | 52.0  | 378   | 2      | US-08-634-224-13  |    | Sequence 13, Appl |
| 17     | 13    | 52.0  | 378   | 2      | US-08-634-224-15  |    | Sequence 15, Appl |
| 18     | 13    | 52.0  | 378   | 2      | US-08-634-400-13  |    | Sequence 13, Appl |
| 19     | 13    | 52.0  | 378   | 2      | US-08-634-400-15  |    | Sequence 15, Appl |
| 20     | 13    | 52.0  | 378   | 2      | US-08-635-878-13  |    | Sequence 13, Appl |
| 21     | 13    | 52.0  | 378   | 2      | US-08-635-878-15  |    | Sequence 15, Appl |
| 22     | 13    | 52.0  | 378   | 2      | US-08-770-057-13  |    | Sequence 13, Appl |
| 23     | 13    | 52.0  | 378   | 2      | US-08-770-057-15  |    | Sequence 15, Appl |
| 24     | 13    | 52.0  | 378   | 4      | US-09-335-697B-13 |    | Sequence 13, Appl |
| 25     | 13    | 52.0  | 378   | 4      | US-09-335-697B-15 |    | Sequence 15, Appl |
| 26     | 13    | 52.0  | 423   | 2      | US-08-345-321-9   |    | Sequence 9, Appl  |
| 27     | 13    | 52.0  | 546   | 3      | US-08-545-809A-26 |    | Sequence 26, Appl |

|      |    |      |      |   |                   |                   |
|------|----|------|------|---|-------------------|-------------------|
| C 28 | 13 | 52.0 | 879  | 3 | US-08-714-071-3   | Sequence 3, Appl  |
| C 29 | 13 | 52.0 | 1173 | 3 | US-08-706-216-5   | Sequence 5, Appl  |
| C 30 | 13 | 52.0 | 1428 | 1 | US-08-488-376-17  | Sequence 17, Appl |
| C 31 | 13 | 52.0 | 1428 | 1 | US-08-488-376-19  | Sequence 19, Appl |
| C 32 | 13 | 52.0 | 1428 | 2 | US-08-634-223-13  | Sequence 13, Appl |
| C 33 | 13 | 52.0 | 1428 | 2 | US-08-634-223-15  | Sequence 15, Appl |
| C 34 | 13 | 52.0 | 1428 | 2 | US-08-634-224-17  | Sequence 17, Appl |
| C 35 | 13 | 52.0 | 1428 | 2 | US-08-634-224-19  | Sequence 19, Appl |
| C 36 | 13 | 52.0 | 1428 | 2 | US-08-634-400-17  | Sequence 17, Appl |
| C 37 | 13 | 52.0 | 1428 | 2 | US-08-634-400-19  | Sequence 19, Appl |
| C 38 | 13 | 52.0 | 1428 | 2 | US-08-635-878-17  | Sequence 17, Appl |
| C 39 | 13 | 52.0 | 1428 | 2 | US-08-635-878-19  | Sequence 19, Appl |
| C 40 | 13 | 52.0 | 1428 | 2 | US-08-770-057-17  | Sequence 17, Appl |
| C 41 | 13 | 52.0 | 1428 | 2 | US-08-770-057-19  | Sequence 19, Appl |
| C 42 | 13 | 52.0 | 1428 | 4 | US-09-335-697B-17 | Sequence 17, Appl |
| C 43 | 13 | 52.0 | 1428 | 4 | US-09-335-697B-19 | Sequence 19, Appl |
| C 44 | 13 | 52.0 | 2830 | 2 | US-09-010-928B-1  | Sequence 1, Appl  |
| C 45 | 13 | 52.0 | 4533 | 3 | US-08-726-214-5   | Sequence 5, Appl  |

## ALIGNMENTS

RESULT 1  
US-08-401-068-7  
; Sequence 7, Application US/08401068  
; Patent No. 5859335  
; GENERAL INFORMATION:  
; APPLICANT: Patton, David  
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PC-DOS/MS-DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/401, 068  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/351, 970  
; FILING DATE: 08-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8614  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1041 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1038  
; IDENTIFICATION METHOD:  
; OTHER INFORMATION: /product- "Biotin synthase"  
; OTHER INFORMATION: /evidence- EXPERIMENTAL  
; US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 5.5e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgaatcctcgtcaggtgcaggtcagc 25  
|||||  
Db 105 CGATCCTCGTCAGGTGCAGGTGACG 129

RESULT 2  
US-08-846-338-7  
; Sequence 7, Application US/08846338  
; Patent No. 5869719  
; GENERAL INFORMATION:  
; APPLICANT: Patton, David  
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5869719artis Corporation  
; STREET: 520 White Plains Road, P.O. Box 2005  
; CITY: Tarrytown  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/846,338  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1041 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEICAL: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1038  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: /product= "biotin synthase"  
; OTHER INFORMATION: /evidence= EXPERIMENTAL  
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 5.5e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgaatcctcgtcaggtgcaggtcagc 25  
|||||  
Db 105 CGATCCTCGTCAGGTGCAGGTGACG 129

RESULT 3  
US-08-411-768B-1  
; Sequence 1, Application US/08411768B  
; Patent No. 6083712  
; GENERAL INFORMATION:  
; APPLICANT: Olwen Birch  
; APPLICANT: Johann Brass  
; APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw  
; TITLE OF INVENTION: Biotechnological Method  
; TITLE OF INVENTION: of Producing Biotin  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/411,768B  
; FILING DATE: 31-March-95  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 3124/92  
; FILING DATE: 02-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 2134/93  
; FILING DATE: 15-JUL-1993  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5872 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: DSM498  
; IMMEDIATE SOURCE:  
; CLONE: PB030A-15/9  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 117..1157  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: /codon\_start= 117  
; OTHER INFORMATION: /product= "Biotin synthase"  
; OTHER INFORMATION: /evidence= EXPERIMENTAL  
; OTHER INFORMATION: /gene= "bioB"  
; OTHER INFORMATION: /number= 1  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2295..3050  
; OTHER INFORMATION: /codon\_start= 2295  
; OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"  
; OTHER INFORMATION: /product= "protein"  
; OTHER INFORMATION: /gene= "bioC"  
; OTHER INFORMATION: /number= 3  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3750..5039  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: /codon\_start= 3750  
; OTHER INFORMATION: /EC\_number= 2.6.1.62  
; OTHER INFORMATION: /product= "DAPA synthase"  
; OTHER INFORMATION: /evidence= EXPERIMENTAL  
; OTHER INFORMATION: /gene= "bioA"  
; OTHER INFORMATION: /number= 5  
; OTHER INFORMATION: /standard\_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate aminotransf." ;  
; OTHER INFORMATION: ;  
; FEATURE:  
; NAME/KEY: CDS



LOCATION: 5098..5574  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /codon\_start= 5098  
OTHER INFORMATION: /function= "unknown", involved in biotin synthesis"  
OTHER INFORMATION: /product= "protein"  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /gene= "ORF1"  
OTHER INFORMATION: /number= 6  
FEATURE:  
NAME/KEY: -10-signal  
LOCATION: 45..49  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /standard\_name= "promoter plac"  
FEATURE:  
NAME/KEY: -35-signal  
LOCATION: 23..28  
OTHER INFORMATION: /standard\_name= "promoter plac"  
FEATURE:  
NAME/KEY: RBS  
LOCATION: 105..119  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /standard\_name= "biob RBS no.9"  
FEATURE:  
NAME/KEY: RBS  
LOCATION: 2284..2297  
OTHER INFORMATION: /standard\_name= "bioc RBS"  
FEATURE:  
NAME/KEY: RBS  
LOCATION: 3742..3752  
OTHER INFORMATION: /standard\_name= "bioa RBS"  
FEATURE:  
NAME/KEY: RBS  
LOCATION: 5088..5100  
OTHER INFORMATION: /standard\_name= "ORF1 RBS"  
FEATURE:  
NAME/KEY: terminator  
LOCATION: 5583..5644  
OTHER INFORMATION: /standard\_name= "rho-independent"  
OTHER INFORMATION: /transcriptional terminator"  
FEATURE:  
NAME/KEY: stem\_loop  
LOCATION: 5583..5605  
FEATURE:  
NAME/KEY: promoter  
LOCATION: 1..96  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /function= "promoter plac"  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
PUBLICATION INFORMATION:  
DOCUMENT NUMBER: WO 87/01391 B1  
FILING DATE: 26-AUG-1986  
PUBLICATION DATE: 07-APR-1993  
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;  
Best Local Similarity 100.0%; Pred. NO. 4.6e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctgcagtcagtcagtcagc 25  
|||||  
Db 221 CGATCCTGCTGAGTGCAGTGCAGC 245

RESULT 4  
US-08-411-768B-6  
; Sequence 6, Application US/08411768B  
; Patent No. 6083712  
; GENERAL INFORMATION:  
; APPLICANT: Olwen Birch  
; APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann  
APPLICANT: Nicholas Shaw  
TITLE OF INVENTION: Biotechnological Method  
TITLE OF INVENTION: of Producing Biotin  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect  
SOFTWARE: Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/411,768B  
FILING DATE: 31-March-95  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 3124/92  
FILING DATE: 02-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 2134/93  
FILING DATE: 15-JUL-1993  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5872 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
STRAIN: DSM498  
IMMEDIATE SOURCE:  
CLONE: PBO30A15-9  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1154..2308  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /codon\_start= 1154  
OTHER INFORMATION: /EC\_number= 2.3.1.47  
OTHER INFORMATION: /product= "KAPA synthase"  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /gene= "bioF"  
OTHER INFORMATION: /number= 2  
OTHER INFORMATION: /standard\_name= "8-Amino-7-oxononanoate synthase"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3043..3753  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /codon\_start= 3043  
OTHER INFORMATION: /EC\_number= 6.3.3.3  
OTHER INFORMATION: /product= "DTB synthase"  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /gene= "bioD"  
OTHER INFORMATION: /number= 4  
OTHER INFORMATION: /standard\_name= "Dethiobiotin synthase"  
FEATURE:  
NAME/KEY: RBS  
LOCATION: 1141..1156  
OTHER INFORMATION: /standard\_name= "bioF RBS"  
FEATURE:  
NAME/KEY: RBS  
LOCATION: 3030..3045  
OTHER INFORMATION: /standard\_name= "biob RBS"  
PUBLICATION INFORMATION:  
DOCUMENT NUMBER: WO 87/01391 B1

FILING DATE: 26-AUG-1986  
PUBLICATION DATE: 07-APR-1993  
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;  
Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatccgcagcagtcagtcagc 25  
|||||  
Db 221 CGATCCTCGTCAGTCAGTCAGC 245

RESULT 5  
US-08-545-809A-5/C  
Sequence 5, Application US/08545809A  
Patent No. 6096878  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Matsuda, Fumihiko  
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME  
NUMBER OF SEQUENCES: 145  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 613 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL TYPE: human lymphoblast  
CELL LINE: GM1  
US-08-545-809A-5

Query Match 56.0%; Score 14; DB 3; Length 613;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcagtcagc 25  
|||||  
Db 283 AGGTGCAGTCAGC 270

RESULT 6  
US-08-545-809A-10/C  
Sequence 10, Application US/08545809A  
Patent No. 6096878  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Matsuda, Fumihiko  
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME  
NUMBER OF SEQUENCES: 145  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 630 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL TYPE: human lymphoblast  
CELL LINE: GM1  
US-08-545-809A-10

Query Match 56.0%; Score 14; DB 3; Length 630;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcagtcagc 25  
|||||  
Db 280 AGGTGCAGTCAGC 267

RESULT 7  
US-09-023-228B-96  
Sequence 96, Application US/09023228B  
Patent No. 6140490  
GENERAL INFORMATION:  
APPLICANT: BIESSECKER, GREGORY  
APPLICANT: GOLD, LARRY  
TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF  
TITLE OF INVENTION: COMPLEMENT SYSTEM PROTEINS  
NUMBER OF SEQUENCES: 157  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Swanson & Bratschun, L.L.C.  
STREET: 8400 E. Prentice Place #200  
CITY: Denver

STATE: Colorado  
COUNTRY: US  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,228B  
FILING DATE: 12-FEBRUARY-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/011739  
FILING DATE: 30 JAN 1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/595,335  
FILING DATE: 1 FEB 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Barry J. Swanson, Esq.  
REGISTRATION NUMBER: 33,215  
REFERENCE/DOCKET NUMBER: NEX50/CIP  
TELEPHONE: (303) 793-3333  
TELEFAX: (303) 793-3433  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA  
FEATURE:  
OTHER INFORMATION: All pyrimidines are 2'-F modified  
US-09-023-228B-96

Query Match 52.0%; Score 13; DB 3; Length 66;  
Best Local Similarity 76.9%; Pred. No. 97;  
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgaatcctgcagcag 13  
|||||:|||||  
DB 41 CGAUCGCGCAGCAG 53

RESULT 8  
US-08-483-636-59/C  
Sequence 59, Application US/08483636  
Patent No. 5914110  
GENERAL INFORMATION:  
APPLICANT: Holmes, Stephen D.  
APPLICANT: Gross, Mitchell S.  
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in  
TITLE OF INVENTION: Treatment of IL4 Mediated Disorders  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corp./Corporate  
ADDRESSEE: Intellectual Property  
STREET: P.O. Box 1539 / UW2220  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,636  
FILING DATE:  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/117366  
FILING DATE: 07-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/136783  
FILING DATE: 14-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US/94/10308  
FILING DATE: 07-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Sutton, Jeffrey A.  
REGISTRATION NUMBER: 34,028  
REFERENCE/DOCKET NUMBER: P50186-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5024  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-483-636-59

Query Match 52.0%; Score 13; DB 2; Length 123;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcagtcag 24  
|||||:|||||  
DB 70 AGTGCAGTCAGCAG 58

RESULT 9  
US-08-483-632-59/C  
Sequence 59, Application US/08483632  
Patent No. 5928904  
GENERAL INFORMATION:  
APPLICANT: Holmes, Stephen D.  
APPLICANT: Gross, Mitchell S.  
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in  
TITLE OF INVENTION: Treatment of IL4 Mediated Disorders  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corp./Corporate  
ADDRESSEE: Intellectual Property  
STREET: P.O. Box 1539 / UW2220  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,632  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/117366  
FILING DATE: 07-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/136783  
FILING DATE: 14-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US/94/10308  
FILING DATE: 07-SEP-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Sutton, Jeffrey A.  
REGISTRATION NUMBER: 34,028  
REFERENCE/DOCKET NUMBER: P50186-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5024  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-483-632-59

Query Match 52.0%; Score 13; DB 2; Length 123;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcaggtcag 24  
|||||  
Db 70 AGGTGCAGGTGAG 58

RESULT 10  
US-08-483-636-63/c  
Sequence 63, Application US/08483636  
Patent No. 5914110  
GENERAL INFORMATION:  
APPLICANT: Holmes, Stephen D.  
APPLICANT: Gross, Mitchell S.  
APPLICANT: Sylvester, Daniel R.  
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in  
TITLE OF INVENTION: Treatment of IL4 Mediated Disorders  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corp./Corporate  
ADDRESS: Intellectual Property  
STREET: P.O. Box 1539 / UW2220  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,636  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/117366  
FILING DATE: 07-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/136783  
FILING DATE: 14-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US/94/10308  
FILING DATE: 07-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Sutton, Jeffrey A.  
REGISTRATION NUMBER: 34,028  
REFERENCE/DOCKET NUMBER: P50186-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5024  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 363 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-483-636-63

Query Match 52.0%; Score 13; DB 2; Length 363;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcaggtcag 24  
|||||  
Db 70 AGGTGCAGGTGAG 58

RESULT 11  
US-08-483-632-63/c  
Sequence 63, Application US/08483632  
Patent No. 5928904  
GENERAL INFORMATION:  
APPLICANT: Holmes, Stephen D.  
APPLICANT: Gross, Mitchell S.  
APPLICANT: Sylvester, Daniel R.  
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in  
TITLE OF INVENTION: Treatment of IL4 Mediated Disorders  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corp./Corporate  
ADDRESS: Intellectual Property  
STREET: P.O. Box 1539 / UW2220  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,632  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/117366  
FILING DATE: 07-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/136783  
FILING DATE: 14-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US/94/10308  
FILING DATE: 07-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Sutton, Jeffrey A.  
REGISTRATION NUMBER: 34,028  
REFERENCE/DOCKET NUMBER: P50186-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5024  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 363 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-483-632-63

Query Match 52.0%; Score 13; DB 2; Length 363;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtcaggtcagc 24  
|||||  
DB 70 AGGTGACAGTCAGC 58

RESULT 12  
US-08-488-376-13/C

Sequence 13, Application US/08488376  
Patent No. 5811524  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,376  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
US-08-488-376-13

Query Match 52.0%; Score 13; DB 1; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtgcaggtcagc 25  
|||||  
DB 69 GGTGACAGTCAGC 57

## RESULT 13

US-08-488-376-15/C  
Sequence 15, Application US/08488376  
Patent No. 5811524  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen

APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,376  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
US-08-488-376-15

Query Match 52.0%; Score 13; DB 1; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtcaggtcagc 24  
|||||  
DB 70 AGGTGACAGTCAGC 58

## RESULT 14

US-08-634-223-13/C  
Sequence 13, Application US/08634223  
Patent No. 5840298  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States

ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/634,223  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
US-08-634-223-13

Query Match 52.0%; Score 13; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtgcaggtcagc 25  
|||||  
Db 69 ggtgcaggtcagc 57

RESULT 15  
US-08-634-223-15/C  
Sequence 15; Application US/08634223  
Patent No. 5840298  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/634,223  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
US-08-634-223-15

Query Match 52.0%; Score 13; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtcaggtcag 24  
|||||  
Db 70 aggtcaggtcag 58

RESULT 16  
US-08-634-224-13/C  
Sequence 13; Application US/08634224  
Patent No. 5866125  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/634,224  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
US-08-634-224-13

Query Match  
Best Local Similarity 52.0%; Score 13; DB 2; Length 378;  
Matches 13; Conservative 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtcaggtcagc 25  
|||||  
Db 69 GGTGAGGTGAGC 57

RESULT 17  
US-08-634-224-15/c  
Sequence 13, Application US/08634224  
Patent No. 5866125  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/634,224  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378

US-08-634-224-15

Query Match  
Best Local Similarity 52.0%; Score 13; DB 2; Length 378;  
Matches 13; Conservative 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcaggtcag 24  
|||||  
Db 70 AGGTGAGGTGAGC 58

RESULT 18  
US-08-634-400-13/c  
Sequence 13, Application US/08634400  
Patent No. 5939068  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/634,400  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
US-08-634-400-13

Query Match  
Best Local Similarity 52.0%; Score 13; DB 2; Length 378;  
Matches 13; Conservative 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtcaggtcagc 25  
|||||  
Db 69 GGTGAGGTGAGC 57

RESULT 19  
US-08-634-400-15/c  
; Sequence 15, Application US/08634400  
; Patent No. 5939068  
; GENERAL INFORMATION:  
; APPLICANT: BRAMS, Peter  
; APPLICANT: CHAMAT, Soulaïma Salim  
; APPLICANT: PAN, Li-Zhen  
; APPLICANT: WALSH, Edward E.  
; APPLICANT: HEARD, Cheryl Janne  
; APPLICANT: NEWMAN, Roland Anthony  
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/634,400  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/488,376  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Teskin, Robin L.  
; REGISTRATION NUMBER: 35,030  
; REFERENCE/DOCKET NUMBER: 012712-150  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 378 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..378  
; US-08-634-400-15

Query Match 52.0%; Score 13; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcaggtcagc 24  
|||||  
DB 70 AGTTCAGGTGACG 58

RESULT 20  
US-08-635-878-13/c  
; Sequence 13, Application US/08635878  
; Patent No. 5955364  
; GENERAL INFORMATION:  
; APPLICANT: BRAMS, Peter  
; APPLICANT: CHAMAT, Soulaïma Salim  
; APPLICANT: PAN, Li-Zhen

; APPLICANT: WALSH, Edward E.  
; APPLICANT: HEARD, Cheryl Janne  
; APPLICANT: NEWMAN, Roland Anthony  
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/635,878  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/488,376  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Teskin, Robin L.  
; REGISTRATION NUMBER: 35,030  
; REFERENCE/DOCKET NUMBER: 012712-150  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 378 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..378  
; US-08-635-878-13

Query Match 52.0%; Score 13; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtcaggtcagc 25  
|||||  
DB 69 GGTTCAGGTGACG 57

RESULT 21  
US-08-635-878-15/c  
; Sequence 15, Application US/08635878  
; Patent No. 5955364  
; GENERAL INFORMATION:  
; APPLICANT: BRAMS, Peter  
; APPLICANT: CHAMAT, Soulaïma Salim  
; APPLICANT: PAN, Li-Zhen  
; APPLICANT: WALSH, Edward E.  
; APPLICANT: HEARD, Cheryl Janne  
; APPLICANT: NEWMAN, Roland Anthony  
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404



CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/635,878  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
US-08-635-878-15

Query Match 52.0%; Score 13; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcaggtcag 24  
|||||  
DB 70 AGGTGCAGGTGACG 58

RESULT 22  
US-08-770-057-13/c  
Sequence 13, Application US/08770057  
Patent No. 5958765  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/770,057  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
US-08-770-057-13

Query Match 52.0%; Score 13; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcaggtcagc 25  
|||||  
DB 69 GGTGCAGGTGACG 57

RESULT 23  
US-08-770-057-15/c  
Sequence 15, Application US/08770057  
Patent No. 5958765  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/770,057  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
US-08-770-057-15

Query Match 52.0%; Score 13; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 agtgcaggtcag 24  
|||||  
Db 70 AGTGACAGGTGAC 58

RESULT 24  
US-09-335-697B-13/C  
Sequence 13, Application US/09335697B  
Patent No. 6200804  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
CHAMAT, Soulaïma Salim  
PAN, Li-Zhen  
WALSH, Edward E.  
HEARD, Cheryl Janne  
NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/335,697B  
FILING DATE: 06-Jul-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/770,057  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-335-697B-13

Query Match 52.0%; Score 13; DB 4; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcaggtcagc 25  
|||||  
Db 69 GGTGACAGGTGACG 57

RESULT 25  
US-09-335-697B-15/C  
Sequence 15, Application US/09335697B  
Patent No. 6200804  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
CHAMAT, Soulaïma Salim  
PAN, Li-Zhen  
WALSH, Edward E.  
HEARD, Cheryl Janne  
NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/335,697B  
FILING DATE: 06-Jul-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/770,057  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..378  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-335-697B-15

Query Match 52.0%; Score 13; DB 4; Length 378;  
Best Local Similarity 100.0%; Pred. No. 82;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 agtgcagtcag 24  
|||||

Db 70 AGTGCAGGTCAG 58

## RESULT 26

US-08-345-321-9/C

; Sequence 9, Application US/08345321

; Patent No. 5914109

; GENERAL INFORMATION:

; APPLICANT: ZOLLA-PAZNER, Susan

; APPLICANT: GORNY, Miroslav K.

; TITLE OF INVENTION: HETEROHYBRIDOMAS PRODUCING HUMAN

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Broadway and Nelmark

; STREET: 419 Seventh Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/345,321

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/872,675

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Broadway, Roger L.

; REGISTRATION NUMBER: 25,618

; REFERENCE/DOCKET NUMBER: ZOLLA-PAZNER1B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

; TELEFAX: 202-737-3528

; TELEX: 248633

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 423 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..423

US-08-345-321-9

Query Match 52.0%; Score 13; DB 2; Length 423;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 agtgcagtcag 24  
|||||

Db 127 AGTGCAGGTCAG 115

## RESULT 27

US-08-545-809A-26/C

; Sequence 26, Application US/08545809A

; Patent No. 6096878

; GENERAL INFORMATION:

; APPLICANT: Honjo, Tasuku

; APPLICANT: Matsuda, Fumihiko

; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE

; TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME

; NUMBER OF SEQUENCES: 145

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/545,809A

; FILING DATE: 27-MAR-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP93/00603

; FILING DATE: 10-MAY-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Freeman, John W.

; REGISTRATION NUMBER: 29,066

; REFERENCE/DOCKET NUMBER: 06501/004001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-542-5070

; TELEFAX: 617-542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 546 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; CELL TYPE: human lymphoblast

; CELL LINE: CGM1

US-08-545-809A-26

Query Match 52.0%; Score 13; DB 3; Length 546;

Best Local Similarity 100.0%; Pred. No. 79;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcagtcagc 25  
|||||

Db 232 GGTGCAGTCAGC 220

## RESULT 28

US-08-714-071-3/C

; Sequence 3, Application US/08714071

; Patent No. 6136584

; GENERAL INFORMATION:

; APPLICANT: Tsutomu, FUJIMARA

; APPLICANT: Shiro, OKUNO

; APPLICANT: Hisanobu, HIRANO

; APPLICANT: Sadahiro, SHIN

; TITLE OF INVENTION: FK506 BINDING PROTEIN GENE

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas

; STREET: 2100 Pennsylvania Avenue, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: United States

; ZIP: 20037-3202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/714,071  
FILING DATE:  
CLASSIFICATION: 435  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 879 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Human fetal brain CDNA library  
IMMEDIATE SOURCE:  
CLONE: OTK4(6-1)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 70...393  
US-08-714-071-3

Query Match 52.0%; Score 13; DB 3; Length 879;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 ggtgcagtcagc 25  
|||||

Db 303 GGTGCAGTCAGC 291

RESULT 29  
US-08-706-216-5/c  
Sequence 5, Application US/08706216  
Patent No. 6140098  
GENERAL INFORMATION:  
APPLICANT: Balasubramanian, SriRam  
APPLICANT: Ford, John  
APPLICANT: Gorman, Daniel M.  
APPLICANT: Zurawski, Gerard  
TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/706,216  
FILING DATE: 30-AUG-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0613  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196  
TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1173 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1173  
US-08-706-216-5

Query Match 52.0%; Score 13; DB 3; Length 1173;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tcaggtcagtc 22  
|||||

Db 418 TCAGGTCAAGTC 406

RESULT 30  
US-08-488-376-17/c  
Sequence 17, Application US/08488376  
Patent No. 5811524  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,376  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1428 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1428  
US-08-488-376-17

Query Match 52.0%; Score 13; DB 1; Length 1428;

Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcagtcagc 25  
|||||  
DB 126 GGTGCAGTCAGC 114

## RESULT 31

US-08-488-376-19/c  
; Sequence 19, Application US/08488376  
; Patent No. 5811524  
; GENERAL INFORMATION:  
; APPLICANT: BRAMS, Peter  
; APPLICANT: CHAMAT, Soulaïma Salim  
; APPLICANT: PAN, Li-Zhen  
; APPLICANT: WALSH, Edward E.  
; APPLICANT: HEARD, Cheryl Janne  
; APPLICANT: NEWMAN, Roland Anthony  
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22133-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,376  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Teskin, Robin L.  
; REGISTRATION NUMBER: 35,030  
; REFERENCE/DOCKET NUMBER: 012712-150  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1428 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1428  
US-08-488-376-19

Query Match 52.0%; Score 13; DB 1; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcagtcagc 24  
|||||  
DB 127 AGGTGAGTCAGC 115

## RESULT 32

US-08-634-223-17/c  
; Sequence 17, Application US/08634223  
; Patent No. 5840298  
; GENERAL INFORMATION:

APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22133-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/634,223  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1428 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1428  
US-08-634-223-17

Query Match 52.0%; Score 13; DB 2; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcagtcagc 25  
|||||  
DB 126 GGTGCAGTCAGC 114

## RESULT 33

US-08-634-223-19/c  
; Sequence 19, Application US/08634223  
; Patent No. 5840298  
; GENERAL INFORMATION:  
; APPLICANT: BRAMS, Peter  
; APPLICANT: CHAMAT, Soulaïma Salim  
; APPLICANT: PAN, Li-Zhen  
; APPLICANT: WALSH, Edward E.  
; APPLICANT: HEARD, Cheryl Janne  
; APPLICANT: NEWMAN, Roland Anthony  
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
; NUMBER OF SEQUENCES: 19

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: P.O. Box 1404  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/634,223  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/488,376  
;; FILING DATE: 07-JUN-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Teskin, Robin L.  
;; REGISTRATION NUMBER: 35,030  
;; REFERENCE/DOCKET NUMBER: 012712-150  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;; INFORMATION FOR SEQ ID NO: 19:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1428 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..1428  
;;  
US-08-634-223-19

Query Match 52.0%; Score 13; DB 2; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcagtcacg 24  
|||||  
Db 127 AGGTGCAGTCACG 115

RESULT 34  
US-08-634-224-17/c  
; Sequence 17, Application US/08634224  
; Patent No. 5866125  
; GENERAL INFORMATION:  
; APPLICANT: BRAMS, Peter  
; APPLICANT: CHAMAT, Soulaïma Salim  
; APPLICANT: PAN, Li-Zhen  
; APPLICANT: WALSH, Edward E.  
; APPLICANT: HEARD, Cheryl Janne  
; APPLICANT: NEWMAN, Roland Anthony  
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/634,224  
;; FILING DATE:  
;; CLASSIFICATION: 424  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/488,376  
;; FILING DATE: 07-JUN-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Teskin, Robin L.  
;; REGISTRATION NUMBER: 35,030  
;; REFERENCE/DOCKET NUMBER: 012712-150  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;; INFORMATION FOR SEQ ID NO: 17:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1428 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..1428  
;;  
US-08-634-224-17

Query Match 52.0%; Score 13; DB 2; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcagtcacg 25  
|||||  
Db 126 GGTGCAGTCACG 114

RESULT 35  
US-08-634-224-19/c  
; Sequence 19, Application US/08634224  
; Patent No. 5866125  
; GENERAL INFORMATION:  
; APPLICANT: BRAMS, Peter  
; APPLICANT: CHAMAT, Soulaïma Salim  
; APPLICANT: PAN, Li-Zhen  
; APPLICANT: WALSH, Edward E.  
; APPLICANT: HEARD, Cheryl Janne  
; APPLICANT: NEWMAN, Roland Anthony  
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/634,224  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/488,376  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35, 030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1428 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1428  
US-08-634-224-19

Query Match 52.0%; Score 13; DB 2; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcaggtcag 24  
|||||  
DB 127 AGGTGCAGGTGCTCAG 115

RESULT 36  
US-08-634-400-17/C  
Sequence 17, Application US/08634400  
Patent No. 5939068  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/634,400  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1428 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1428  
US-08-634-400-17

Query Match 52.0%; Score 13; DB 2; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcaggtcagc 25  
|||||  
DB 126 GGTGCAGGTGCTCAGC 114

RESULT 37  
US-08-634-400-19/C  
Sequence 19, Application US/08634400  
Patent No. 5939068  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
APPLICANT: CHAMAT, Soulaïma Salim  
APPLICANT: PAN, Li-Zhen  
APPLICANT: WALSH, Edward E.  
APPLICANT: HEARD, Cheryl Janne  
APPLICANT: NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/634,400  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/488,376  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1428 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1428  
US-08-634-400-19

Query Match 52.0%; Score 13; DB 2; Length 1428;

Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcaggtcag 24  
|||||  
Db 127 AGGTGAGGTGAG 115

RESULT 38  
US-08-635-878-17/c

; Sequence 17, Application US/08635878  
; Patent No. 5955364

; GENERAL INFORMATION:

; APPLICANT: BRAMS, Peter

; APPLICANT: CHAMAT, Soulaïma Salim

; APPLICANT: PAN, Li-Zhen

; APPLICANT: WALSH, Edward E.

; APPLICANT: HEARD, Cheryl Janne

; APPLICANT: NEWMAN, Roland Anthony

; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN

; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/635,878

; FILING DATE:

; CLASSIFICATION:

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 08/488,376

; FILING DATE: 07-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Teskin, Robin L.

; REGISTRATION NUMBER: 35,030

; REFERENCE/DOCKET NUMBER: 012712-150

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ. ID NO: 17:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1428 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1428

; US-08-635-878-17

QY 13 ggtcaggtcagc 25  
|||||  
Db 126 GGTGAGGTGAGC 114

RESULT 39  
US-08-635-878-19/c

; Sequence 19, Application US/08635878  
; Patent No. 5955364

; GENERAL INFORMATION:

; APPLICANT: BRAMS, Peter

; APPLICANT: CHAMAT, Soulaïma Salim

; APPLICANT: PAN, Li-Zhen

; APPLICANT: WALSH, Edward E.

; APPLICANT: HEARD, Cheryl Janne

; APPLICANT: NEWMAN, Roland Anthony

; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN

; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/635,878

; FILING DATE:

; CLASSIFICATION:

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 08/488,376

; FILING DATE: 07-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Teskin, Robin L.

; REGISTRATION NUMBER: 35,030

; REFERENCE/DOCKET NUMBER: 012712-150

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ. ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1428 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1428

; US-08-635-878-19

QY 12 aggtcaggtcag 24  
|||||  
Db 127 AGGTGAGGTGAG 115

RESULT 40  
US-08-770-057-17/c

; Sequence 17, Application US/08770057  
; Patent No. 5958765

; GENERAL INFORMATION:

; APPLICANT: BRAMS, Peter

; APPLICANT: CHAMAT, Soulaïma Salim

; APPLICANT: PAN, Li-Zhen

; APPLICANT: WALSH, Edward E.

; APPLICANT: HEARD, Cheryl Janne

; APPLICANT: NEWMAN, Roland Anthony

; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN

Query Match 52.0%; Score 13; DB 2; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



;; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
;; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
;; NUMBER OF SEQUENCES: 19  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: P.O. Box 1404  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/770,057  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/488,376  
;; FILING DATE: 07-JUN-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Teskin, Robin L.  
;; REGISTRATION NUMBER: 35,030  
;; REFERENCE/DOCKET NUMBER: 012712-150  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;; INFORMATION FOR SEQ ID NO: 17:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1428 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..1428  
;;  
;; US-08-770-057-17

Query Match 52.0%; Score 13; DB 2; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtcgaggtcagc 25  
|||||  
DB 126 GGTGCGAGTCTCAGC 114

RESULT 41  
US-08-770-057-19/c  
;; Sequence 19, Application US/08770057  
;; Patent No. 5958765  
;; GENERAL INFORMATION:  
;; APPLICANT: BRAMS, Peter  
;; APPLICANT: CHAMAT, Soulaïma Salim  
;; APPLICANT: PAN, Li-Zhen  
;; APPLICANT: WALSH, Edward E.  
;; APPLICANT: HEARD, Cheryl Janne  
;; APPLICANT: NEWMAN, Roland Anthony  
;; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
;; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
;; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
;; NUMBER OF SEQUENCES: 19  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: P.O. Box 1404  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/770,057  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/488,376  
;; FILING DATE: 07-JUN-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Teskin, Robin L.  
;; REGISTRATION NUMBER: 35,030  
;; REFERENCE/DOCKET NUMBER: 012712-150  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;; INFORMATION FOR SEQ ID NO: 19:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1428 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..1428  
;;  
;; US-08-770-057-19

Query Match 52.0%; Score 13; DB 2; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcgaggtcag 24  
|||||  
DB 127 AGTTCGAGTCTCAG 115

RESULT 42  
US-09-335-697B-17/c  
;; Sequence 17, Application US/09335697B  
;; Patent No. 6200804  
;; GENERAL INFORMATION:  
;; APPLICANT: BRAMS, Peter  
;; APPLICANT: CHAMAT, Soulaïma Salim  
;; APPLICANT: PAN, Li-Zhen  
;; APPLICANT: WALSH, Edward E.  
;; APPLICANT: HEARD, Cheryl Janne  
;; APPLICANT: NEWMAN, Roland Anthony  
;; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
;; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
;; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
;; NUMBER OF SEQUENCES: 19  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: P.O. Box 1404  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/335,697B  
;; FILING DATE: 06-Jul-2000  
;; CLASSIFICATION: <Unknown>  
;; PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/770,057  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1428 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1428  
SEQUENCE DESCRIPTION: SEQ ID NO: 17:  
US-09-335-697B-17

Query Match 52.0%; Score 13; DB 4; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtgcaggtcagc 25  
126 ggtgcaggtcagc 114

RESULT 43  
US-09-335-697B-19/c  
Sequence 19, Application US/09335697B  
Patent No. 6200804  
GENERAL INFORMATION:  
APPLICANT: BRAMS, Peter  
CHAMAT, Soulaïma Salim  
PAN, Li-Zhen  
WALSH, Edward E.  
HEARD, Cheryl Janne  
NEWMAN, Roland Anthony  
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN  
MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND  
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/335,697B  
FILING DATE: 06-Jul-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/770,057  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1428 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1428  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-09-335-697B-19

Query Match 52.0%; Score 13; DB 4; Length 1428;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 agtgcaggtcagc 24  
127 agtgcaggtcagc 115

RESULT 44  
US-09-010-928B-1  
Sequence 1, Application US/09010928B  
Patent No. 5994099  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V  
APPLICANT: Hayashi, Cheryl Y  
TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 GATEHOUSE RD, SUITE 500E  
CITY: FALLS CHURCH  
STATE: VIRGINIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 22042  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/010,928B  
FILING DATE: 22-JAN-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28977  
REFERENCE/DOCKET NUMBER: 1447-109P  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2830 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..2830  
OTHER INFORMATION: /note="Flagelliform DNA sequence  
taken from the 5' region. The putative start codon is at  
position 219"  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 219..2830  
US-09-010-928B-1

Query Match 52.0%; Score 13; DB 2; Length 2830;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcaggtcaggtc 22  
 |||  
 Db 588 TCAGTCAAGTC 600

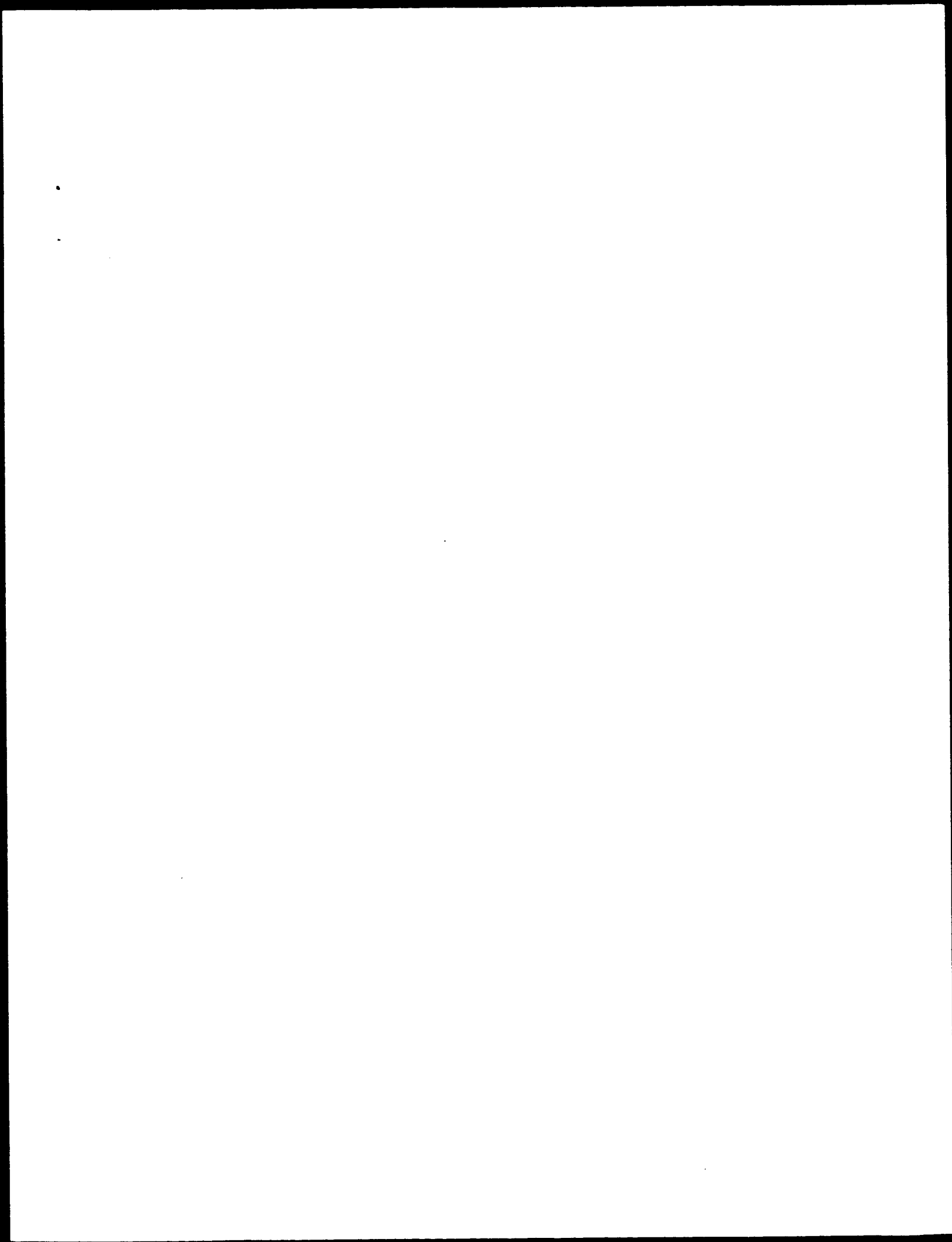
## RESULT 45

US-08-726-214-5  
 ; Sequence 5, Application US/08726214  
 ; Patent No. 6107076  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Wei-Jen  
 ; APPLICANT: Gilman, Alfred G.  
 ; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE  
 ; TITLE OF INVENTION: AND USES THEREFOR  
 ; NUMBER OF SEQUENCES: 31  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Arnold, White & Durkee  
 ; STREET: P.O. Box 4433  
 ; CITY: Houston  
 ; STATE: Texas  
 ; COUNTRY: United States of America  
 ; ZIP: 77210  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/726,214  
 ; FILING DATE: Concurrently Herewith  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/005,498  
 ; FILING DATE: 04-OCT-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Highlander, Steven L.  
 ; REGISTRATION NUMBER: 37,642  
 ; REFERENCE/DOCKET NUMBER: UTSD:450  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (512) 418-3000  
 ; TELEFAX: (512) 474-7577  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4533 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-726-214-5

Query Match 52.0%; Score 13; DB 3; Length 4533;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcaggtcagc 25  
 |||  
 Db 2664 GGTGCAAGTCAGC 2676

Search completed: October 9, 2001, 15:55:42  
 Job time: 13218 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:37 ; Search time 5323.87 Seconds  
(without alignments)  
44.389 Million cell updates/sec

Title: US-09-396-196f-9  
Perfect score: 25  
Sequence: 1 cgatcctcgtcagtcagtcagtcagc 25

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 475250

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.\*

```
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*
48: gb_est48:*
49: gb_est49:*
50: gb_est50:*
51: gb_est51:*
52: gb_est52:*
53: gb_est53:*
54: gb_est54:*
55: gb_est55:*
56: gb_est56:*
57: gb_est57:*
58: gb_est58:*
59: gb_est59:*
60: gb_est60:*
61: gb_est61:*
62: gb_est62:*
63: gb_est63:*
64: gb_est64:*
65: gb_est65:*
66: gb_est66:*
67: gb_est67:*
68: gb_est68:*
69: gb_est69:*
70: gb_est70:*
71: gb_est71:*
72: gb_est72:*
73: gb_est73:*
74: gb_est74:*
75: gb_est75:*
76: gb_est76:*
77: gb_est77:*
78: gb_est78:*
79: gb_est79:*
80: gb_est80:*
81: gb_est81:*
82: gb_est82:*
83: gb_est83:*
84: gb_est84:*
85: gb_est85:*
86: gb_est86:*
87: gb_est87:*
88: gb_est88:*
89: gb_est89:*
90: gb_est90:*
91: gb_est91:*
92: gb_est92:*
93: gb_est93:*
94: gb_est94:*
95: gb_est95:*
96: gb_est96:*
97: gb_est97:*
98: gb_est98:*
99: gb_est99:*
100: gb_est100:*
101: gb_est101:*
102: gb_est102:*
103: gb_est103:*
104: gb_est104:*
105: gb_est105:*
106: gb_est106:*
107: gb_est107:*
108: gb_est108:*
109: gb_est109:*
110: gb_est110:*
111: gb_est111:*
112: gb_est112:*
113: gb_est113:*
114: gb_est114:*
115: gb_est115:*
116: gb_est116:*
```

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

---

```

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| C 1        | 17    | 68.0        | 324    | 149   | BF464374    |
| C 2        | 17    | 68.0        | 413    | 141   | BE862745    |
| C 3        | 17    | 68.0        | 495    | 24    | A1789420    |
| C 4        | 17    | 68.0        | 641    | 243   | A2426129    |
| C 5        | 16    | 64.0        | 163    | 1     | AA065941    |
| C 6        | 16    | 64.0        | 288    | 160   | BB516544    |
| C 7        | 16    | 64.0        | 328    | 190   | W54143      |
| C 8        | 16    | 64.0        | 338    | 190   | W34110      |
| C 9        | 16    | 64.0        | 368    | 2     | AA110497    |
| C 10       | 16    | 64.0        | 370    | 121   | AW822106    |
| C 11       | 16    | 64.0        | 434    | 20    | A1467556    |
| C 12       | 16    | 64.0        | 540    | 122   | AW892602    |
| C 13       | 16    | 64.0        | 553    | 229   | AO530462    |
| C 14       | 16    | 64.0        | 775    | 136   | BE535032    |
| C 15       | 16    | 64.0        | 822    | 77    | BE194624    |
| C 16       | 16    | 64.0        | 897    | 220   | CNS021CM    |
| C 17       | 16    | 64.0        | 976    | 174   | BG172968    |
| C 18       | 16    | 64.0        | 1011   | 132   | AK010837    |
| C 19       | 16    | 64.0        | 1298   | 152   | BG328084    |
| C 20       | 15    | 60.0        | 180    | 10    | AA699872    |
| C 21       | 15    | 60.0        | 264    | 157   | H22139      |
| C 22       | 15    | 60.0        | 279    | 187   | H22139      |
| C 23       | 15    | 60.0        | 293    | 251   | A2877721    |
| C 24       | 15    | 60.0        | 300    | 127   | BB175346    |
| C 25       | 15    | 60.0        | 306    | 173   | BG100182    |
| C 26       | 15    | 60.0        | 345    | 31    | AV633196    |
| C 27       | 15    | 60.0        | 395    | 148   | BF444068    |
| C 28       | 15    | 60.0        | 396    | 116   | AA486040    |
| C 29       | 15    | 60.0        | 407    | 1     | AA062258    |
| C 30       | 15    | 60.0        | 407    | 151   | BF599674    |
| C 31       | 15    | 60.0        | 408    | 224   | AO085001    |
| C 32       | 15    | 60.0        | 409    | 151   | BF653430    |
| C 33       | 15    | 60.0        | 431    | 122   | AW925187    |
| C 34       | 15    | 60.0        | 457    | 151   | BF604067    |
| C 35       | 15    | 60.0        | 477    | 225   | AO921336    |
| C 36       | 15    | 60.0        | 492    | 140   | BE808281    |
| C 37       | 15    | 60.0        | 496    | 241   | AZ261270    |
| C 38       | 15    | 60.0        | 496    | 150   | BF652819    |
| C 39       | 15    | 60.0        | 519    | 167   | BF442332    |
| C 40       | 15    | 60.0        | 543    | 239   | AZ168021    |
| C 41       | 15    | 60.0        | 607    | 251   | AZ873604    |
| C 42       | 15    | 60.0        | 624    | 243   | AZ418216    |
| C 43       | 15    | 60.0        | 639    | 114   | AW342762    |
| C 44       | 15    | 60.0        | 658    | 156   | C98409      |
| C 45       | 15    | 60.0        | 668    | 108   | AU165518    |

## ALIGNMENTS

RESULT 1  
BF464374/c  
LOCUS  
DEFINITION  
UI-M-CG0p-bog-f-09-0-UI.s1 NIH\_BMAP\_Ret4\_s2  
VERSION  
ACCESSION  
KEYWORDS  
SOURCE  
ORGANISM

324 bp mRNA  
EST  
Mus musculus cDNA clone  
BF464374  
GI:11533557  
EST.  
house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
1 (bases 1 to 324)  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)

## MEDLINE

97044477  
Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: MEST@mail.nih.gov

Oligo-dt track not found. Not I site shown in beginning of sequence  
is likely internal to the message. cDNA library preparation: M.B.  
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
clones from RESEARCH GENETICS. It should be noted that Bento Soares  
is generating a small number of additional specialized  
non-redundant arrays of BMAP cDNAs whose availability will be  
considered under appropriate and limited collaborative arrangements  
The following repetitive elements were found in this cDNA sequence:  
102-169, >(GAA)n#Simple\_repeat  
Seq primer: M13 Forward  
POLYA-No.

## FEATURES

Location/Qualifiers  
1..324  
/organism="Mus musculus"  
/strain="C57Bl/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-CG0p-bog-f-09-0-UI"  
/clone\_1lb="NIH\_BMAP\_Ret4\_S2"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; The  
NIH\_BMAP\_Ret4\_S2 library is a subtracted library,  
ultimately derived from mouse retina tissue libraries at  
various stages of development. For a detailed description  
of the library from which this clone was derived, please  
visit our web site at brainest.eng.uiowa.edu.  
TAG\_Sp0=None found"

## BASE COUNT

91 a 84 c 89 g 60 t

## Query Match

68.0%; Score 17; DB 149; Length 324;  
Best Local Similarity 100.0%; Pred. No. 14;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## Qy

8 cgtcaggtgcaggtcag 24  
|||||

## Db

324 CGTCAGGTGCAGGTCA 308

RESULT 2  
BE862745/c  
LOCUS  
DEFINITION  
UI-M-BH0-ajf-b-11-0-UI.r1 NIH\_BMAP\_M.S1 Mus musculus cDNA clone  
VERSION  
ACCESSION  
KEYWORDS  
SOURCE  
ORGANISM

413 bp mRNA  
EST  
Mus musculus  
BE862745  
GI:10382015  
EST.  
house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
1 (bases 1 to 413)  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)

## MEDLINE

97044477  
Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: MEST@mail.nih.gov





```

/clone.lib="Mouse 10kb plasmid UUGCM library"
/sex="Male"
/lab.host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42m; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (q147321141gb/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      156 a      184 c      183 g      118 t
ORIGIN

```

```

Query Match      68.0%; Score 17; DB 243; Length 641;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      8 cgtcaggtcaggtcag 24
        |||||
Db      212 CGTCAGGTGACAGTCAG 196

```

```

RESULT 5
LOCUS   AA065941 163 bp mRNA EST 03-FEB-1997
DEFINITION m151a01.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
ACCESSION AA065941
VERSION   AA065941.1 GI:1562644
KEYWORDS EST.
SOURCE    house mouse.
ORGANISM Mus musculus

```

```

REFERENCE 1 (bases 1 to 163)
AUTHORS  Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
          Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
          Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
          Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
          Waterston, R.

```

```

TITLE    The WashU-HHMI Mouse EST Project
JOURNAL  Unpublished (1996)
COMMENT  Contact: Marra M/Mouse EST Project
          WashU-HHMI Mouse EST Project
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
          Tel: 314 286 1800
          Fax: 314 286 1810

```

```

FEATURES
Source    Email: mouseest@wustl.edu
          This clone is available royalty-free through LLNL; contact the
          IMGE Consortium (info@image.llnl.gov) for further information.
          Seq primer: -28m13 rev1 ET from Amersham
          High quality sequence stop: 144.
          location/Qualifiers
          1..163
          /organism="Mus musculus"
          /strain="Inbred CD-1"
          /db_xref="taxon:10090"
          /clone="IMAGE:515496"

```

```

/clone.lib="Stratagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: testis; Vector: pBluescript SK-; Site:1;
       EcoRI; Site:2; XhoI; Cloned unidirectionally. Primer:
       Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
       -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
       sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'"
BASE COUNT      23 a      56 c      49 g      35 t
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      9 gtcaggtcaggtcag 24
        |||||
Db      16 GTCAGGTGACAGTCAG 1

```

```

RESULT 6
LOCUS   BS16544/c 288 bp mRNA EST 28-JUL-2000
DEFINITION BS16544 RIKEN full-length enriched, 16 days neonate heart Mus
          musculus cDNA clone DB300163j11.3, similar to M76601 Mouse alpha
          cardiac myosin heavy chain mRNA, mRNA sequence.
ACCESSION BS16544
VERSION   BS16544.1 GI:9568002
KEYWORDS EST.
SOURCE    house mouse.
ORGANISM Mus musculus

```

```

REFERENCE 1 (bases 1 to 288)
AUTHORS  Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci,
          P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
          Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
          Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
          Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusabe, M.,
          Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
          Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
          Y., Shigemoto, Y., Shinagawa, A., Shirai, T., Sogabe, Y., Sugahara, Y.,
          Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,
          T., Tsunoda, Y., Watanishi, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
          Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
          M., Muramatsu, M. and Hayashizaki, Y.

```

```

TITLE    RIKEN Mouse ESTs (Kono, H., et al.)
JOURNAL  Unpublished (2000)
COMMENT  Contact: Yoshihide Hayashizaki
          Genome Exploration Research Group, Life Science Tsukuba Center,
          Genome Science Laboratory
          The Institute of Physical and Chemical Research (RIKEN), Genomic
          Sciences Center
          3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
          Tel: +81-298-36-9013
          Fax: +81-298-36-9098
          Email: genome-res@rctc.riken.go.jp,
          URL: http://genome.rtc.riken.go.jp/

```

```

          Email: mouse-res@rctc.riken.go.jp,
          URL: http://genome.rtc.riken.go.jp/
          N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
          Thermolabile and thermocatalytic of thermolabile enzymes by
          trehalose and its application for the synthesis of full length
          cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
          Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
          Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
          Y. and Hayashizaki, Y.
          Automated filtration-based high-throughput plasmid preparation
          system. Genome Res. 9 (5), 463-470 (1999)
          Carninci, P. and Hayashizaki, Y.
          High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

```

19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.  
Location/Qualifiers  
1 200

|            |      |       |      |      |
|------------|------|-------|------|------|
| BASE COUNT | 83 a | 104 c | 57 g | 44 t |
| ORIGIN     |      |       |      |      |

```

query Match      64.0%; Score 16; DB 160; Length 288;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0;

```

|    |     |                  |     |
|----|-----|------------------|-----|
| Qy | 8   | cytcaggtgcaggtca | 23  |
|    |     |                  |     |
| Db | 188 | CGTCAGGTCAGGTCA  | 173 |

| RESULT     | 7                  |              |        |           |      |                   |
|------------|--------------------|--------------|--------|-----------|------|-------------------|
| W54143     |                    |              |        |           |      | 03-JUN-1996       |
| LOCUS      |                    |              |        |           |      |                   |
| DEFINITION | W54143             | 328          | bp     | MRNA      |      |                   |
|            | md13c03.f1         | Seares mouse | embryo | NBME13.5  | 14.5 | Mus musculus cDNA |
|            | Clone IMAGE:368260 | 5',          | RNA    | sequence. |      |                   |

REFERENCE  
1 (bases 1 to 328)  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.

|         |                                    |
|---------|------------------------------------|
| TITLE   | The WashU-HHMI Mouse EST Project   |
| JOURNAL | Unpublished (1996)                 |
| COMMENT | Contact: Marra M/Mouse EST Project |

## FEATURES

### Location/Qualifiers

Source

```

source
1. .328
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="368260"
/clone_id="Soares mouse embryo NBM13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dp total fetus"
/lab_host="DH10B"
/notes="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCGAATTTTTTTTTTTTTTTT
3'], on equal amounts of mRNA from 2 13.5dp and 2
14.5dp embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo. "

```

|                       |                 |               |           |             |
|-----------------------|-----------------|---------------|-----------|-------------|
| Query Match           | 64.0%;          | Score 16;     | DB 190;   | Length 328; |
| Best Local Similarity | 100.0%;         | Pred. No. 48; |           |             |
| Matches 16;           | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0;     |

```
QY      10 tcaggtcaggtcagc 25
          |||
Db      86 TCAGGTGAGGTCAGC 101
```

|            |  |        |      |     |             |
|------------|--|--------|------|-----|-------------|
| RESULT     | 8  |        |      |     |             |
| W41110     |  |        |      |     |             |
| LOCUS      | W41110   | 338 bp | mRNA | EST | 13-MAY-1996 |
| DEFINITION | mc57c606.r1 Soares mouse embryo NDbME13.5 14.5 Mus musculus cDNA clone IMAGE:352618 5', mRNA sequence. |        |      |     |             |

|           |                              |
|-----------|------------------------------|
| ACCESSION | W34110                       |
| VERSION   | W34110.1                     |
| KEYWORDS  | GI:1316081                   |
| SOURCE    | EST.                         |
| ORGANISM  | house mouse,<br>Mus musculus |

REFERENCE 1 (Pages 1 to 338)  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

|         |                                    |
|---------|------------------------------------|
| TITLE   | The WashU-HIMI Mouse EST Project   |
| JOURNAL | Unpublished (1996)                 |
| COMMENT | Contact: Marra M/Mouse EST Project |

| FEATURES | source                  | location/qualifiers |
|----------|-------------------------|---------------------|
|          | 1. .338                 |                     |
|          | /organism="Mus musculus |                     |
|          | /strain="C57BL/6J"      |                     |
|          | /db_xref="taxon:10090"  |                     |

```

/clone="IMAGE:352618"
/clone_lib="Soares mouse embryo NBMEL3.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAAGTGGAGCGCGCGGAAATTTTATTTTATTTTATTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pRT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT      86 a      65 c      84 g      103 t
ORIGIN

Query Match      64.0%; Score 16; DB 190; Length 338;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      10 tcaggtgcagtcagc 25
Db      82 TCAGGTGCAGTCAGC 97
|||||
|||||

RESULT 9
A1110497/c 368 bp mRNA EST 03-FEB-1997
LOCUS m162e03.r1 Stratagene mouse testis (#937308) Mus musculus CDNA
DEFINITION clone IMAGE:516604 5', mRNA sequence.
ACCESSION A1110497
VERSION A1110497.1 GI:1662274
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 368)
Marras,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Maria M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:310452
Seq primer: -28mj3 rev1 ET from Amersham.

FEATURES
Source location/Qualifiers
1..368
/organism="Mus musculus"
/strain="Inbred CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:516604"
/clone_lib="Stratagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"

```

```

/note="Organ: testis; Vector: plusescript SK-; Site-1:
EcoRI; Site-2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTATTTTATTTTATTTTATTTT 3'"
BASE COUNT      75 a      115 c      106 g      72 t
ORIGIN

Query Match      64.0%; Score 16; DB 2; Length 368;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      9 gtcaggtgcagtcag 24
Db      21 GTCAGGTGCAGTCAG 6
|||||
|||||

RESULT 10
AM822106 370 bp mRNA EST 17-MAY-2000
LOCUS uq13c09.x1 Ren Stubbs mouse thymus Mus musculus CDNA clone
DEFINITION IMAGE:2802352 3', mRNA sequence.
ACCESSION AM822106
VERSION AM822106.1 GI:7915183
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 370)
Marras,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurr,R., Ritter
,E., Kohn,S., Shu,T., Jackson,T., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Other ESTs: uq13c09.y1
Contact: Maria M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1041876
Seq primer: Primer name ambiguous
High quality sequence stop: 303.

FEATURES
Source location/Qualifiers
1..370
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:2802352"
/clone_lib="Ren Stubbs mouse thymus"
/sex="mixed"
/dev_stage="3 weeks"
/lab_host="DH10B"
/note="Organ: thymus; Vector: pRT73D-Pac; Site-1: NotI;
Site-2: PacI; 1st strand cDNA was primed with an oligo(dT)
primer: double-stranded cDNA was ligated using 5' linker
ggcgcat and 3' linker aactgaagactatatt. Library is
size-selected >2.5 kb and average insert size is 3.5 kb.
Clones were arrayed from primary plating, non-amplified.
Library constructed by X. Ren and L. Stubbs (Lawrence
Livermore National Laboratory and DOE Joint Genome
Institute, 7000 East Ave, L-453, Livermore, CA 94550)."
BASE COUNT      125 a      57 c      75 g      113 t
ORIGIN

```

Query Match 64.0% Score 16; DB 121; Length 370;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 gtcaggtcaggtcag 24  
 |||  
 Db 53 gtcaggtcaggtcag 68

RESULT 11  
 A1467556 430 bp mRNA EST 09-MAR-1999  
 LOCUS ve36g09.x1 Soares\_mammary\_gland\_NbMMG Mus musculus CDNA clone  
 DEFINITION IMAGE:820288 3', mRNA sequence.  
 ACCESSION A1467556 GI:4320893  
 VERSION A1467556.1 GI:4320893  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 430)  
 NC1-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
 MGI:488568  
 This clone was previously sequenced on the 5' end only, this new  
 data is from the 3' end  
 High quality sequence stop: 406.

FEATURES  
 source  
 Location/Qualifiers  
 1..430  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_image="820288"  
 /clone\_lib="Soares\_mammary\_gland\_NbMMG"  
 /sex="male"  
 /tissue\_type="mammary gland"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Organ: mammary gland; Vector: pRT3D-pac (Pharmacia  
 ) with a modified polylinker. Site\_1: Not I; Site\_2: Eco  
 RI; 1st strand cDNA was primed with a Not I - Oligo(dT)  
 primer (5'  
 TGTTCACCAATCTCAAGTGGAGCGGCCGGAATGTTTTTTTTTTTTTTTTTTTT  
 T 3'); double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pRT3D vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
 constructed and normalized by Bento Soares and M.Fatima  
 Bonafido."

BASE COUNT 136 a 102 c 85 g 107 t

Query Match 64.0% Score 16; DB 20; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcaggtcaggtcag 25  
 |||  
 Db 259 TCAGGTGAGGTGAGC 244

RESULT 12  
 AM892602 544 bp mRNA EST 24-MAY-2000  
 LOCUS CM3-NN0004-100300-111-f07 NN0004 Homo sapiens CDNA, mRNA sequence.  
 DEFINITION

ACCESSION AM892602  
 VERSION AM892602.1 GI:8056807  
 EST.  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 544)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the PAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-cm3-NN0004-100>  
 300-111-f07&t3=2000-03-10&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 543.

FEATURES  
 source  
 Location/Qualifiers  
 1..544  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NN0004"  
 /dev\_stage="Adult"  
 /note="Organ: nervous\_normal; Vector: puc18. Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

BASE COUNT 127 a 152 c 160 g 105 t

Query Match 64.0% Score 16; DB 122; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 gtcaggtcaggtcag 24  
 |||  
 Db 383 gtcaggtcaggtcag 398

RESULT 13  
 A0530462 553 bp DNA GSS 18-MAY-1999  
 LOCUS RPCI-11-369F11.TV RPCI-11 Homo sapiens genomic clone RPCI-11-369F11  
 DEFINITION RPCI-11-369F11.TV RPCI-11 Homo sapiens genomic clone RPCI-11-369F11  
 , DNA sequence.  
 ACCESSION A0530462  
 VERSION A0530462.1 GI:4842505  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 553)  
 Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter,  
 J.C.

|                       |  |                  |   |
|-----------------------|--|------------------|---|
| JOURNAL               | TITLE  |                  | Use of BAC End Sequences from Library RPCT-11 for Sequence-Ready Map Building |
| COMMENT               | Other-GSSS: RPCT-11-369F11-TV  |                  | Unpublished (1997)  |
|                       | Contact: Shaying Zhao, William Nierman, Mark Adams   |                  |   |
|                       | Department of Eukaryotic Genomics  |                  |   |
|                       | The Institute for Genomic Research   |                  |   |
|                       | 9712 Medical Center Dr., Rockville, MD 20850   |                  |   |
|                       | Tel: 301 838 0200  |                  |   |
|                       | Fax: 301 838 0208  |                  |   |
|                       | Email: dhoe@tigr.org   |                  |   |
|                       | Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pieter de Jong (pieter@jeong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html. |                  |   |
| FEATURES              | Seq primer: SP6  | Class: BAC ends. |   |
| source                | location/Qualifiers  |                  |   |
|                       | 1..553   |                  |   |
|                       | /organism="Homo sapiens"   |                  |   |
|                       | /db_xref="GDB:7641442"   |                  |   |
|                       | /db_xref="taxon:9606"  |                  |   |
|                       | /clone="RPCT-11-369F11"  |                  |   |
|                       | /clone_lib="RPCT-11"   |                  |   |
|                       | /sex="Male"  |                  |   |
|                       | /cell_type="Lymphocytes"   |                  |   |
|                       | /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCT11 Human Male BAC library"   |                  |   |
| BASE COUNT            | 108 a  | 165 c            | 157 g   |
| ORIGIN                |  |                  | 123 t   |
| Query Match           | 64.0%;   | Score 16;        | DB 229; Length 553;   |
| Best Local Similarity | 100.0%;  | Pred. No. 48;    |   |
| Matches               | 16;  | Conservative     | 0; Mismatches 0; Indels 0; Gaps 0;  |
| QY                    | 9 gtcaggtgcaggtcag 24  |                  |   |
| Db                    | 204 GTCAGTGCAGGTCAG 189  |                  |   |
| RESULT 14             |  |                  |   |
| LOCUS                 | BE535032   | 775 bp           | mRNA  |
| DEFINITION            | 601233227F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3596935 5',  | EST              | 09-AUG-2000   |
| ACCESSION             | BE535032   |                  |   |
| VERSION               | BE535032   |                  |   |
| KEYWORDS              | EST  |                  |   |
| SOURCE                | house mouse.   |                  |   |
| ORGANISM              | Mus musculus   |                  |   |
| REFERENCE             | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  |                  |   |
| AUTHORS               | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.   |                  |   |
| TITLE                 | 1 (bases 1 to 775)   |                  |   |
| JOURNAL               | NIH-MGC http://mgc.nci.nih.gov/  |                  |   |
| COMMENT               | Unpublished (1999)   |                  |   |
|                       | Contact: Robert Strausberg, Ph.D.  |                  |   |
|                       | Email: cgabbs-r@mail.nih.gov   |                  |   |
|                       | Tissue procurement: Jeffrey Green M.D.   |                  |   |
|                       | cDNA library Preparation: Life Technologies, Inc.  |                  |   |
|                       | cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  |                  |   |
|                       | DNA sequencing by: Incyte Genomics, Inc.   |                  |   |
|                       | Clone distribution: MGC clone distribution information can be  |                  |   |
|                       | found through the I.M.A.G.E. Consortium/LLNL at:   |                  |   |
|                       | http://image.llnl.gov  |                  |   |
|                       | Plate: LHAM8775 row: f column: 08  |                  |   |
|                       | High quality sequence stop: 225.   |                  |   |
|                       | Location/Qualifiers  |                  |   |
|                       | 1..775   |                  |   |
| FEATURES              | /organism="Mus musculus"   |                  |   |
| source                |  |                  |   |

|                       |  |                  |   |
|-----------------------|--|------------------|---|
| JOURNAL               | TITLE  |                  | Use of BAC End Sequences from Library RPCT-11 for Sequence-Ready Map Building |
| COMMENT               | Other-GSSS: RPCT-11-369F11-TV  |                  | Unpublished (1997)  |
|                       | Contact: Shaying Zhao, William Nierman, Mark Adams   |                  |   |
|                       | Department of Eukaryotic Genomics  |                  |   |
|                       | The Institute for Genomic Research   |                  |   |
|                       | 9712 Medical Center Dr., Rockville, MD 20850   |                  |   |
|                       | Tel: 301 838 0200  |                  |   |
|                       | Fax: 301 838 0208  |                  |   |
|                       | Email: dhoe@tigr.org   |                  |   |
|                       | Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pieter de Jong (pieter@jeong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html. |                  |   |
| FEATURES              | Seq primer: SP6  | Class: BAC ends. |   |
| source                | location/Qualifiers  |                  |   |
|                       | 1..553   |                  |   |
|                       | /organism="Homo sapiens"   |                  |   |
|                       | /db_xref="GDB:7641442"   |                  |   |
|                       | /db_xref="taxon:9606"  |                  |   |
|                       | /clone="RPCT-11-369F11"  |                  |   |
|                       | /clone_lib="RPCT-11"   |                  |   |
|                       | /sex="Male"  |                  |   |
|                       | /cell_type="Lymphocytes"   |                  |   |
|                       | /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCT11 Human Male BAC library"   |                  |   |
| BASE COUNT            | 108 a  | 165 c            | 157 g   |
| ORIGIN                |  |                  | 123 t   |
| Query Match           | 64.0%;   | Score 16;        | DB 229; Length 553;   |
| Best Local Similarity | 100.0%;  | Pred. No. 48;    |   |
| Matches               | 16;  | Conservative     | 0; Mismatches 0; Indels 0; Gaps 0;  |
| QY                    | 9 gtcaggtgcaggtcag 24  |                  |   |
| Db                    | 204 GTCAGTGCAGGTCAG 189  |                  |   |
| RESULT 15             |  |                  |   |
| LOCUS                 | BE194624   | 822 bp           | cDNA  |
| DEFINITION            | 30-JUL-2000 (Rel. 64, Created)   |                  |   |
| ACCESSION             | BE194624   |                  |   |
| VERSION               | BE194624   |                  |   |
| KEYWORDS              | EST  |                  |   |
| SOURCE                | Hordeum vulgare (barley)   |                  |   |
| ORGANISM              | Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  |                  |   |
| REFERENCE             | Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.   |                  |   |
| TITLE                 | 1 (bases 1 to 822)   |                  |   |
| JOURNAL               | Wing R., Close T.J., Kleinholz A., Wise R., Begum D., Frisch D., Yu Y.,  |                  |   |
| COMMENT               | Anderson H., Dale J., Henry D., Kernodle S., Palmer M., Rambo T.,  |                  |   |
|                       | Saski C., Schwartzbeck J., Simmons J., Choi D.W., Main D., Wood T.;  |                  |   |
|                       | "Development of a genetically and physically anchored EST resource for   |                  |   |
|                       | barley genomics";  |                  |   |
|                       | Unpublished.   |                  |   |
|                       | Contact: Wing RA   |                  |   |
|                       | Clemson University Genomics Institute  |                  |   |
|                       | 100 Jordan Hall, Clemson, SC 29634, USA  |                  |   |
|                       | Tel: 864 656 7288  |                  |   |
|                       | Fax: 864 656 4293  |                  |   |
|                       | Email: twing@clemson.edu   |                  |   |
|                       | Seq primer: ATTATACCCCTACTAAGG   |                  |   |
|                       | High quality sequence start: 225   |                  |   |
|                       | High quality sequence stop: 345.   |                  |   |
|                       | Location/Qualifiers  |                  |   |
|                       | 1..822   |                  |   |
|                       | /db_xref="taxon:4513"  |                  |   |
|                       | /db_xref="ESTLIB:341"  |                  |   |
|                       | /note="Vector: LambdaZAP, site_1: EcoRI; site_2: XhoI"   |                  |   |
|                       | /organism="Hordeum vulgare"  | </               |   |

```

FT      /clone="HVSMEH0086E11f"
FT      /clone_lib="Hordium vulgare 5-45 DAP spike EST library
FT      HVCDDN0009 (5 to 45 DAP)"
FT      /tissue_type="5-45 DAP Spike"
FT      /lab_host="SOLR"
XX      SO      Sequence 822 bp; 276 A; 160 C; 184 G; 200 T; 2 other;

Query Match
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 gtcaggtgcaggtcag 24
        |||
Db      362 GTCAGGTGCAGGTGACG 347

RESULT 16
CNS021CM 897 bp DNA GSS 12-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
DEFINITION 225C12 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL176719 GI:7814776
VERSION AL176719.1 GI:7814776
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetradon.
REFERENCE 1 (bases 1 to 897)
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 897)
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
COMMENT Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 897)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetradon.
FEATURES
source
1..897
/organism="Tetradon nigroviridis"
/db_xref="taxon:99863"
/clone="225C12"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG225BB06LPI-end : T7"
BASE COUNT 118 a 284 c 294 g 182 t 19 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 16; DB 220; Length 897;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 gtcaggtgcaggtcag 24
        |||
Db      16 GTCAGGTGCAGGTGACG 31

```

```

RESULT 17
CNS021CM 976 bp mRNA EST 06-FEB-2001
LOCUS BG172968/c BG172968/c
DEFINITION BG172968 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4459758 5',
mRNA sequence.
ACCESSION BG172968
VERSION BG172968.1 GI:12679671
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 976)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM10260 row: e column: 07
High quality sequence stop: 514.
FEATURES
source
1..976
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4459758"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site1: SalI;
Site2: NotI; Cloned unidirectionally. Primer: oligo dr.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 255 a 250 c 163 g 308 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 16; DB 174; Length 976;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 gtcaggtgcaggtcag 24
        |||
Db      168 GTCAGGTGCAGGTGACG 153

RESULT 18
CNS021CM 1011 bp mRNA HTC 08-FEB-2001
LOCUS AK010837 AK010837/c
DEFINITION Mus musculus ES cells cDNA, RIKEN full-length enriched library,
clone:2410198H06, full insert sequence.
ACCESSION AK010837
VERSION AK010837.1 GI:12846558
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA,
clone_lib=RIKEN full-length enriched mouse cDNA library
clone:2410198H06.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning

```

|  |   |
|--|---|
| JOURNAL<br>REFERENCE<br>AUTHORS            | Methods Enzymol. 303, 19-44 (1999)  |
| TITLE                                      | 2 (sites)<br>Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.<br>Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes<br>Genome Res. 10 (10), 1617-1630 (2000)   |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS | 3 (sites)<br>Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Kimuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsunoto,H., Sakauechi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Iwata,M., Ohara,E., Matsubara,S., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsubara,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.<br>RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer<br>Genome Res. 10 (11), 1757-1771 (2000)  |
| TITLE                                      | 4 (sites)<br>The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.<br>Functional annotation of a full-length mouse cDNA collection<br>Nature 409, 685-690 (2001)   |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS | 5 (bases 1 to 1011)<br>Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.<br>Direct Submission<br>Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.res.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216) |
| COMMENT                                    | Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details.  |
| FEATURES                                   | Location/Qualifiers<br>1..1011<br>/organism="Mus musculus"<br>/strain="C57BL/6J"<br>/db_xref="taxon:10090"<br>/db_xref="MGD:MGI:1910427"<br>/db_xref="MGD:MGI:1924031"<br>/clone="2410198H06"<br>/cell_type="ES cells"<br>/clone_idb="RIKEN full-length enriched mouse cDNA library"  |
| BASE COUNT                                 | 255 a 227 c 180 g 349 t   |
| ORIGIN                                     |   |
| Query Match                                | 64.0%: Score 16; DB 192; Length 1011;   |
| Best Local Similarity                      | 100.0%: Pred. No. 47;   |
| Matches                                    | 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;   |
| OY   | 9 gtcaggtgcagctaac 24<br>   |
| Db   | 230 GTCAGGTGCAGCTAAG 275  |
| RESULT 19                                  |   |
| BG328084/c                                 | 1298 bp mRNA EST 27-FEB-2001  |
| LOCUS                                      |   |

|                          |  |  |  |  |
|--------------------------|--|--|--|--|
| DEFINITION               | 602427134F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546497 5', mRNA sequence.  |  |  |  |
| ACCESSION                | BG328084   |  |  |  |
| VERSION                  | BG328084.1 GI:13134522   |  |  |  |
| KEYWORDS                 | EST.   |  |  |  |
| SOURCE                   | human.   |  |  |  |
| ORGANISM                 | Homo sapiens   |  |  |  |
| REFERENCE                | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  |  |  |  |
| AUTHORS                  | 1 (bases 1 to 1298)  |  |  |  |
| TITLE                    | NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .  |  |  |  |
| JOURNAL                  | National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)  |  |  |  |
| COMMENT                  | Contact: Robert Strusberg, Ph.D.<br>Email: <a href="mailto:cgabbs-r@mail.nih.gov">cgabbs-r@mail.nih.gov</a><br>Tissue Procurement: ATCC<br>cDNA Library Preparation: Ling Hong/Rubin Laboratory<br>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)<br>DNA Sequencing by: NIH Intramural Sequencing Center<br>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a><br>Plate: LLCM123 row: c column: 10<br>High quality sequence stop: 511.<br>Location/Qualifiers<br>1..1298<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/clone="IMAGE:4546497"<br>/clone_1ib="NIH_MGC_15"<br>/tissue_type="adenocarcinoma cell line"<br>/lab_host="DH10B (phage-resistant)"<br>/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Size selected >500bp for average insert. Size 1.8kb. Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)." |  |  |  |
| BASE COUNT               | 316 a 498 c 357 g 127 t  |  |  |  |
| ORIGIN                   |  |  |  |  |
| Query Match              | 64.0%; Score 16; DB 152; Length 1298;  |  |  |  |
| Best Local Similarity    | 100.0%; Pred. No. 46;  |  |  |  |
| Matches 16; Conservative | 0; Mismatches 0; Indels 0; Gaps 0;   |  |  |  |
| Oy                       | 7 tccatcaggtgcagctc 22<br>   |  |  |  |
| Db                       | 984 TCGTCAGGTGCAGCTC 969   |  |  |  |
| RESULT 20                | AA699872 180 bp mRNA EST 19-DEC-1997   |  |  |  |
| LOCUS                    | Z9J1F04.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:461311 3', mRNA sequence.  |  |  |  |
| ACCESSION                | AA699872   |  |  |  |
| VERSION                  | AA699872.1 GI:2702835  |  |  |  |
| KEYWORDS                 | EST.   |  |  |  |
| SOURCE                   | human.   |  |  |  |
| ORGANISM                 | Homo sapiens   |  |  |  |
| REFERENCE                | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  |  |  |  |
| AUTHORS                  | 1 (bases 1 to 180)   |  |  |  |
| TITLE                    | Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacey, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theisling, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.  |  |  |  |
| JOURNAL                  | WashU-NCI human EST Project  |  |  |  |
| COMMENT                  | Unpublished (1997)<br>Contact: Wilson RK   |  |  |  |

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu

This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 144.

## FEATURES

source

```
1..180
/organism="Homo sapiens"
/db_xref="GDB:375234"
/db_xref="taxon:9606"
/clone="IMAGE:461311"
/clone.lib="Soares_fetal_liver_spleen_1NPLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INPLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
AACTGAGAGATTATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
```

BASE COUNT  
ORIGIN

47 a 57 c 36 g 40 t

## Query Match

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcagtcagtcagc 25

Db 138 CAGGTGCAGGTCAAC 124

## RESULT 21

H22139/c

LOCUS H22139 264 bp mRNA EST 06-JUL-1995  
DEFINITION Y138a03.s1 Soares breast 3NbHst Homo sapiens cDNA clone  
IMAGE:160492 3', mRNA sequence.

ACCESSION H22139  
VERSION H22139.1 GI:890834

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 264)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman  
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston  
R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)

TITLE JOURNAL  
COMMENT

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu

Insert Size: 685  
High quality sequence stops: 241  
Source: IMAGE Consortium, LNLN

This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 685 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 241.

## FEATURES

source

```
1..264
/organism="Homo sapiens"
/db_xref="GDB:574535"
/db_xref="taxon:9606"
/clone="IMAGE:160492"
/clone.lib="Soares breast 3NbHst"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGAGCGAGCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of a modified pT73 vector (Pharmacia).
Library went through one round of normalization to a Cot =
20. Library constructed by Bento Soares and M.Fatima
Bonaldo."
```

BASE COUNT  
ORIGIN

87 a 47 c 50 g 73 t 7 others

## Query Match

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcaagtcagtcagtcag 24

Db 215 TCAGGTGCAGGTCAAC 201

## RESULT 22

R50061/c

LOCUS R50061 279 bp mRNA EST 18-MAY-1995  
DEFINITION YJ59c10.s1 Soares breast 2NbHst Homo sapiens cDNA clone  
IMAGE:153042 3', mRNA sequence.

ACCESSION R50061  
VERSION R50061.1 GI:811963

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 279)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman  
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston  
R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)

TITLE JOURNAL  
COMMENT

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu

Insert Size: 681  
High quality sequence stops: 257 Source: IMAGE Consortium, LNLN

This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 681 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 257.

## FEATURES

source

```
1..279
/organism="Homo sapiens"
/db_xref="GDB:565317"
/db_xref="taxon:9606"
```



```

re derived from the mouse BAC library RPc1-23. For BAC
availability, please contact Pieter de Jong
(email: cho.org). Clones may be purchased from BACPAC
s (http://www.choi.org/bacpac/orderingframe.htm). BAC end
95 row: 1 column: 20
ter: 17
BAC ends:
Location/Qualifiers
1. 293
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPc1-23-195120"
/clone_lib="RPc1-23"
/sex="Female"
/lab_host="RDH108"
/notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1;
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

```

```

FEATURES
source      Location/Qualifiers
1. .300    /organism="Mus musculus"

```

```

/db_xref="taxon:10090"
/clone="A230056E21"
/clone_lib="RIKEN full-length enriched, adult male
hypothalamus"
/sex="male"
/tissue_type="hypothalamus"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site 1: Salt; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGACAGACAGATCCAAAGACCTCTTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGACAGACATCTCGAGTTAATTAATATCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

```

```

BASE COUNT      66 a      81 c      58 g      95 t
ORIGIN

```

```

Query Match      60.0%; Score 15; DB 127; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      10 tcaggtcaggtcag 24
        |||||
Db       72 TCAGGTGTCAGGTGTCAG 58

```

```

RESULT 25      306 bp      mRNA      EST      29-JAN-2001
BG100182/c      LOCUS
DEFINITION      ux94904.y1 McCarrey Eddy type B spermatogonia Mus musculus cDNA
                clone IMAGE:3656262 5', mRNA sequence.
ACCESSION      BG100182
VERSION        BG100182.1 GI:12595499
KEYWORDS
SOURCE
ORGANISM

```

```

REFERENCE
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 306)
Marrin, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person
, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuit, R., Rutter
, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R., and Wilson, R.
The Mashu-NCI Mouse EST Project 1999
Unpublished (1999)

```

```

TITLE
JOURNAL
COMMENT      Contact: Marrin M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through INLNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1418566
Seq primer: Primer name ambiguous.
Location/Qualifiers
1..306
/organism="Mus musculus"
/strain="CD-1"
/db_xref="taxon:10090"

```

FEATURES

source

```

/clone="IMAGE:3656262"
/clone_lib="McCarrey Eddy type B spermatogonia"
/sex="male"
/tissue_type="type B spermatogonia, pooled from multiple
mice"
/dev_stage="8 day"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pBluescript SK+ (Stratagene
); Site 1: XhoI; Site 2: EcoRI; cDNA oligo dt-primed
[5'-(GA)10-ACCTGCTCGAGTTTCTTTTCTTNN 3'] and directionally
cloned using 5' linkers 5'-AATTCGACGACG-3' and
5'-CTGCTGCGC-3'. Size selection of >400bp material gives
average insert size ranging from 1-2 kb. Library was mass
excised (from lambda-unizap-XR) and resulting
single-stranded phagemids were prepped and transformed
into DH10B. Library contains 96% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63417."

```

```

BASE COUNT      62 a      82 c      93 g      69 t
ORIGIN

```

```

Query Match      60.0%; Score 15; DB 173; Length 306;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      9 gtcaggtcaggtca 23
        |||||
Db       95 CTCAGGTGTCAGGTCA 81

```

```

RESULT 26      345 bp      mRNA      EST      15-DEC-2000
AV633196/c      LOCUS
DEFINITION      AV633196 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
                cDNA clone HCO17h07_r 5', mRNA sequence.
ACCESSION      AV633196
VERSION        AV633196.1 GI:10776516
KEYWORDS
SOURCE
ORGANISM

```

```

REFERENCE
AUTHORS      Chlamydomonas reinhardtii.
1 (bases 1 to 345)
Asamizu, E., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y., and Tabata, S.
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
DNA Res. 7 (5), 305-307 (2000)
20539644

```

```

TITLE
JOURNAL
COMMENT      Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..345
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCO17h07_r"
/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"

```

FEATURES

source

```

BASE COUNT      62 a      105 c      98 g      79 t      1 others
ORIGIN

```

Query Match 60.0%; Score 15; DB 31; Length 345;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caggtgcaggtcagc 25  
 |||  
 DB 173 CAGGTGCAGGTGACG 159

RESULT 27  
 LOCUS BF444068 395 bp mRNA EST 01-DEC-2000  
 DEFINITION 261798 MARC 2Pig Sus scrofa cDNA 5', mRNA sequence.  
 ACCESSION BF444068  
 VERSION BF444068.1 GI:11504160  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 1 (bases 1 to 395)  
 Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.,  
 and Keele,J.W.  
 Design and use of two pooled tissue normalized cDNA libraries for  
 EST discovery in swine  
 Unpublished (2000)  
 CONTACT: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called and alt trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -mismatch 12 options.  
 PCR Primers  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCACGTCAGCAGC  
 Plate: 96 row: D column: 1  
 Seq primer: ATTAGTGCAGCTATAG.  
 Location/Qualifiers  
 1..395  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone\_lib="MARC 2Pig"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
 library made from pooled tissue from testis, ovary,  
 endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 118 a 66 c 143 g 68 t

ORIGIN

Query Match 60.0%; Score 15; DB 148; Length 395;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caggtgcaggtcagc 25  
 |||  
 DB 309 CAGGTGCAGGTGACG 323

RESULT 28  
 LOCUS AA486040 396 bp mRNA EST 10-JUL-2000  
 DEFINITION 69428 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.  
 ACCESSION AA486040  
 VERSION AA486040.1 GI:7056146  
 KEYWORDS EST.

SOURCE COW.  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 396)  
 Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,  
 Bennett,G.A., Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid  
 ,W.W. and Keele,J.W.  
 Design and use of four pooled tissue normalized cDNA libraries for  
 EST discovery in cattle  
 Unpublished (2000)  
 CONTACT: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 20  
 and -mismatch 12 options.  
 PCR Primers  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCACGTCAGCAGC  
 Plate: 45 row: I column: 7  
 Seq primer: ATTAGTGCAGCTATAG.  
 Location/Qualifiers  
 1..396  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="MARC 4BOV"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
 library made from pooled tissue from day 20 and day 40  
 embryos."

BASE COUNT 70 a 134 c 127 g 65 t

ORIGIN

Query Match 60.0%; Score 15; DB 116; Length 396;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caggtgcaggtcagc 25  
 |||  
 DB 221 CAGGTGCAGGTGACG 207

RESULT 29  
 LOCUS AA062258 407 bp mRNA EST 03-FEB-1997  
 DEFINITION m155908.r1 Striatagene mouse testis (#937308) Mus musculus cDNA  
 clone IMAGE:515966 5', mRNA sequence.  
 ACCESSION AA062258  
 VERSION AA062258.1 GI:1556057  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 407)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 CONTACT: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:309814  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 387.

## FEATURES

source

Location/Qualifiers

1..407

/organism="Mus musculus"

/strain="Inbred CD-1"

/db\_xref="taxon:10090"

/clone="IMAGE:515966"

/clone\_lib="Scratogene mouse testis (#937308)"

/sex="males"

/tissue\_type="testis"

/dev\_stage="10-12 week old"

/lab\_host="SOLR (kanamycin resistant)"

/note="Organ: testis; Vector: pBluescript SK-; Site\_1:

EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:

0.190 dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;

-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor

sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

## BASE COUNT

88 a 117 c 123 g 79 t

ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 15; DB 1; Length 407;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 gtcagtcagtcagtc 23

Db 15 CTCAGTCAGTCAGTC 1

## RESULT 30

BF599674

407 bp mRNA

EST

13-DEC-2000

LOCUS

263546 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION

BF599674

VERSION

BF599674.1 GI:11696393

KEYWORDS

EST.

ORGANISM

Bos taurus

REFERENCE

Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,

AUTHORS

Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid

TITLE

Design and use of four pooled tissue normalized cDNA libraries for

JOURNAL

EST discovery in cattle

COMMENT

Unpublished (2000)

CONTACT

Contact: Smith TPL

ADDRESS

USDA, ARS, US Meat Animal Research Center

PO BOX

166, Clay Center, NE 68933-0166, USA

TEL

402 762 4366

FAX

402 762 4390

EMAIL

smitht@mail.marc.usda.gov

NOTES

Single pass sequencing. Bases called and alt. trimmed with phred

SOURCE

1..407

/organism="Bos taurus"

/db\_xref="taxon:9913"  
 /clone\_lib="MARC 3BOV"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV SPOR6; Site\_1: XbaI; Site\_2: XhoI;  
 Library made from pooled tissue from marrow, alveolar  
 macrophage, ovary, fetal semitendinosus muscle, and fetal  
 longissimus muscle."

## BASE COUNT

69 a 139 c 135 g 64 t

ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 15; DB 151; Length 407;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caggtcagtcagtc 25

Db 93 CAGGTGCGAGTCAGC 79

## RESULT 31

A0085001/c

408 bp DNA

GSS

26-AUG-1998

LOCUS

HS\_2269\_A2\_B10\_MR CIT Approved Human Genomic Sperm Library D Homo

ACCESSION

A0085001

VERSION

A0085001.1 GI:3454218

KEYWORDS

GSS.

ORGANISM

human.

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE

Mahairas G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

JOURNAL

Mahairas G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

MEDLINE

1 (bases 1 to 408)

COMMENT

Sequence-tagged connectors: A sequence approach to mapping and

CONTACT

scanning the human genome

PROC.

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

ADDRESS

99380589

CONTACT

Contact: Mahairas GG, Wallace JC, Hood L

UNIVERSITY

High Throughput Sequencing Center

TITLE

University of Washington

ADDRESS

401 Queen Anne Avenue North, Seattle, WA 98109, USA

TEL

Tel: (206) 616-3618

FAX

Fax: (206) 616-3887

EMAIL

Email: jwallace@u.washington.edu

SEQUENCE

Sequence Tagged Connector

PLATE

Plate: 2269 row: C column: 20

CLASS

Class: BAC ends

HIGH QUALITY

High quality sequence stop: 408.

FEATURES

Location/Qualifiers

source

1..408

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="plate-2269 Col-20 Row-C"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in

E-Coli DH10B"

BASE COUNT

104 a 111 c 78 g 115 t

ORIGIN

QY 11 caggtcagtcagtc 25

Db 396 CAGGTGCGAGTCAGC 382

Query Match

Best Local Similarity 100.0%; Score 15; DB 224; Length 408;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 32  
BF653430/c 409 bp mRNA EST 20-DEC-2000  
LOCUS  
DEFINITION 277257 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BF653430  
VERSION BF653430.1 GI:11918562  
KEYWORDS EST.  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 409)  
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,  
Bennett,G.A., Fahrnkruug,S.C., Freking,B.A., Rohrer,G.A., Laegreid  
,W.W. and Keeler,J.W.  
Design and use of four pooled tissue normalized cDNA libraries for  
EST discovery in cattle  
Unpublished (2000)  
JOURNAL  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACACGCTATGACCAT  
BACKWARD: GTTTCAGTCAGCAGC  
Plate: 69 row: P column: 10  
Seq primer: ATTTCAGTCAGCTATAG.  
Location/Qualifiers  
1. 409  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 3BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from marrow, alveolar  
macrophage, ovary, fetal semitendinosus muscle, and fetal  
longissimus muscle."  
BASE COUNT 80 a 138 c 129 g 62 t  
ORIGIN

Query Match 60.0%; Score 15; DB 151; Length 409;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caggtgcaggtcagc 25  
|||||  
Db 255 CAGGTGCAGGTGACG 241

RESULT 33  
AM925187 431 bp mRNA EST 19-JUL-2000  
LOCUS  
DEFINITION WSI\_76\_E02.bl\_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,  
mRNA sequence.  
ACCESSION AM925187  
VERSION AM925187.1 GI:8091013  
KEYWORDS EST.  
SOURCE  
ORGANISM Sorghum  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoidae; Andropogoneae; Sorghum.  
1 (bases 1 to 431)  
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt

TITLE  
JOURNAL  
COMMENT An EST database from Sorghum: water-stressed plants  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.  
Seq primer: JEN REV  
High quality sequence stop: 339  
POLYA-No.  
Location/Qualifiers  
1. 431  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_lib="Water-stressed 1 (WS1)"  
/note="Organ: Mix of 5-week old plants on days 7 & 8 after  
water was withheld; Vector: lambda Zap; Site\_1: XhoI;  
Site\_2: EcoRI; The library was made from polyA RNA in the  
cloning vector lambda Zap II. Clones to be sequenced were  
prepared by mass excision."  
BASE COUNT 98 a 135 c 106 g 92 t  
ORIGIN

Query Match 60.0%; Score 15; DB 122; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatccgcagtcagtc 16  
|||||  
Db 419 GATCCTCTCAGGTG 405

RESULT 34  
BF604067 457 bp mRNA EST 13-DEC-2000  
LOCUS  
DEFINITION 269755 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BF604067  
VERSION BF604067.1 GI:11702307  
KEYWORDS EST.  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 457)  
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,  
Bennett,G.A., Fahrnkruug,S.C., Freking,B.A., Rohrer,G.A., Laegreid  
,W.W. and Keeler,J.W.  
Design and use of four pooled tissue normalized cDNA libraries for  
EST discovery in cattle  
Unpublished (2000)  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACACGCTATGACCAT  
BACKWARD: GTTTCAGTCAGCAGC  
Plate: 50 row: I column: 16  
Seq primer: ATTTCAGTCAGCTATAG.  
Location/Qualifiers

```

source
1. .457
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_1lb="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT      82 a      152 c      142 g      80 t      1 others
ORIGIN

Query Match      60.0%; Score 15; DB 151; Length 457;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      11 caggtgcaggtcagc 25
|||||
Db      215 CAGGTGCAGGTGACG 201

RESULT 35
A0921326/c      477 bp      DNA      GSS      21-DEC-1999
LOCUS      A0921326
DEFINITION      RPCI-23-273p15. TV RPCI-23 Mus musculus genomic clone RPCI-23-273p15
, DNA sequence.
ACCESSION      A0921326
VERSION      A0921326.1 GI:6610329
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 477)
Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Aklnret
, B., Levins, M., Megann, S., Jsegeye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@edj.org.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac/ends/mouse/bac\_end\_intro.html
Plate: 273 row: P column: 15
Seq primer: T7
Class: BAC ends.
FEATURES
source
1. .477
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1lb="RPCI-23-273p15"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

```

```

BASE COUNT      101 a      107 c      125 g      144 t
ORIGIN

Query Match      60.0%; Score 15; DB 235; Length 477;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      9 gtcaagtcaggtcagc 23
|||||
Db      68 GTCAGTGCAGGTGCA 54

RESULT 36
BE808281/c      492 bp      mRNA      EST      20-SEP-2000
LOCUS      BE808281
DEFINITION      213447 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION      BE808281
VERSION      BE808281.1 GI:10239393
KEYWORDS      EST.
SOURCE      cow.
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 492)
Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M.,
Bennett, G.A., Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Laegreid
, W.W. and Keele, J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
Contact: Smith TP
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
V0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACACAT
BACKWARD: GTTTCAGTACGACGACG
Plate: 68 row: P column: 2
Seq primer: ATTAGGTGACACTATAG.
FEATURES
source
1. .492
Location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_1lb="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT      96 a      161 c      149 g      85 t      1 others
ORIGIN

Query Match      60.0%; Score 15; DB 140; Length 492;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      11 caggtgcaggtcagc 25
|||||
Db      263 CAGGTGCAGGTGACG 249

RESULT 37
A2261270/c      495 bp      DNA      GSS      26-JUL-2000
LOCUS      A2261270

```

```

DEFINITION  RPCI-23-44018.TV RPCI-23 Mus musculus genomic clone RPCI-23-44018,
ACCESSION   DNA sequence.
VERSION     AZ261270
KEYWORDS    AZ261270.1 GI:9469499
SOURCE      GSS.
ORGANISM    house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 495)
AUTHORS     Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akinet
            ,B., Levis,M., McGinn,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-23
            Unpublished (1999)
            Other GSSs: RPCI-23-44018.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 440 row: 1 column: 8
            Seq primer: 17
            Class: BAC ends.
FEATURES
source
    1..495
    Location/Qualifiers
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone_lib="RPCI-23-44018"
        /clone_1lb="RPCI-23"
        /sex="Female"
        /lab_host="DH10B"
        /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
        EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
        brain genomic DNA was isolated and partially digested
        with a combination of EcoRI and EcoRI Methylase. Size
        selected DNA was cloned into the pBACe3.6 vector at the
        EcoRI sites. The ligation products were transformed into
        DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT  106 a 110 c 132 g 147 t
ORIGIN
Query Match          60.0%; Score 15; DB 240; Length 495;
Best Local Similarity 100.0%; Pred. NO. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 gtcaggtcaggtca 23
    ||||||||||||
Db 66 GTCAGGTGACGTCA 52

```

```

REFERENCE 1 (bases 1 to 496)
AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keefe,J.W.
            Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
            Unpublished (2000)
            Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smithemil@marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -mismatch 12 options.
            PCR primers
            FORWARD: AGGAACAGCTATGACCAT
            BACKWARD: GTTTCGCCATCAGCAGC
            Plate: 67 row: C column: 19
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES
source
    1..496
    Location/Qualifiers
        /organism="Bos taurus"
        /db_xref="taxon:9913"
        /clone_1lb="MARC 3BOV"
        /tissue_type="pooled"
        /lab_host="DH10B"
        /note="Vector: PCMV SPORTE; Site_1: XbaI; Site_2: XhoI;
        Library made from pooled tissue from marrow, alveolar
        macrophage, ovary, fetal semitendinosus muscle, and fetal
        longissimus muscle."
BASE COUNT  100 a 164 c 148 g 84 t
ORIGIN
Query Match          60.0%; Score 15; DB 151; Length 496;
Best Local Similarity 100.0%; Pred. NO. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caggtcaggtcagc 25
    ||||||||||||
Db 290 CAGGTGACGTGACG 276

```

```

RESULT 39
BE442332
LOCUS      BE442332      519 bp      mRNA      EST      25-JUL-2000
DEFINITION Chlamydomonas reinhardtii CC-2290, normalized, Lambda Zap II
ACCESSION  BE442332
VERSION    BE442332.1 GI:9441847
KEYWORDS   EST.
SOURCE     Chlamydomonas reinhardtii.
ORGANISM   Chlamydomonas reinhardtii.
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE  1 (bases 1 to 519)
AUTHORS    Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Sillflow,C., Stern,D. and Surzycki,R.
            Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants: Project phase 2
            Unpublished (2000)
            Contact: Elizabeth H. Harris
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlmy@duke.edu.
JOURNAL    Unpublished (2000)
COMMENT    Project phase 2

```

```

FEATURES
source
    1..519
    Location/Qualifiers

```

/organism="Chlamydomonas reinhardtii"  
 /strain="CC-2290 wild type mt- S1 D2"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-2290, normalized, lambda Zap  
 II"  
 /note="Vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; This library was constructed by John Davies and  
 Jeffrey McDermott. RNA was isolated from strain CC-2290  
 (Minnesota isolate of C. reinhardtii) grown to mid-log  
 phase in TAP (acetate containing) medium in the light.  
 PolyA mRNA was purified, and cDNA was synthesized and  
 directionally cloned into lambda Zap II (Stratagene) in  
 the EcoRI (5') and XhoI (3') sites. pBluescript II SK-  
 plasmids were excised from the lambda Zap clones by  
 superinfection with ExAssist (Stratagene) phage. The  
 library was normalized using method 4 described in Bonaldo  
 et al (1996) Genome Research 6: 791-806."

|                       |         |                    |         |               |
|-----------------------|---------|--------------------|---------|---------------|
| Query Match           | 60.0%;  | Score 15;          | DB 167; | Length 519;   |
| Best Local Similarity | 100.0%; | Pred. No. 1.7e+02; |         |               |
| Matches               | 15;     | Conservative       | 0;      | Mismatches 0; |
|                       |         |                    | Indels  | 0;            |
|                       |         |                    | Gaps    | 0             |

|    |    |                 |    |
|----|----|-----------------|----|
| QY | 4  | tcctcgtcaggtgca | 18 |
|    |    |                 |    |
| Db | 72 | TCCTCGTCAGGTGCA | 86 |

| RESULT     | 40  |
|------------|---|
| AL168021/c |   |
| LOCUS      |   |
| DEFINITION | AZ168021 543 bp DNA GSS 29-AUG-2000                                 |
|            | SP 0103 BI F05-T7A Strongylocentrotus purpuratus, purple sea urchin |
|            | , sperm genomic BAC library Strongylocentrotus purpuratus genomic   |
|            | clone Plate=103 Col=9 Row=L, DNA sequence.                          |

|           |            |
|-----------|------------|
| ACCESSION | AZ168021   |
| VERSION   | AZ168021.1 |
|           | GI:8338389 |

SOURCE  
ORGANISM

REFERENCE  
AUTHORS

11

JOURNAL  
MEDLINE  
20402566  
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 150  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: [acameron@caltech.edu](mailto:acameron@caltech.edu)  
Plate: 103 row: L column: 9  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 543.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source   | 1. .543             |

```

/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="plate103 Col-9 Row-L"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli
DH10B"

```

| BASE COUNT | 149 a | 116 c | 116 g | 157 t | 5 others |
|------------|-------|-------|-------|-------|----------|
| ORIGIN     |       |       |       |       |          |

|                       |                 |                    |           |             |
|-----------------------|-----------------|--------------------|-----------|-------------|
| Query Match           | 60.0%;          | Score 15;          | DB 239;   | Length 543; |
| Best Local Similarity | 100.0%;         | Pred. No. 1.7e+02; |           |             |
| Matches 15;           | Conservative 0; | Mismatches 0;      | Indels 0; | Gaps 0;     |

|    |     |                  |     |
|----|-----|------------------|-----|
| Qy | 3   | atcctcgtaggtgc   | 17  |
|    |     |                  |     |
| Db | 115 | ATCCTCGTCAAGGTGC | 101 |

|            |   |
|------------|---|
| RESULT     | 41  |
| AZ873604/c |   |
| LOCUS      | AZ873604 607 bp DNA GSS   |
| DEFINITION | ZM0187J24 Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGCZM0187J24 F, DNA sequence. |
|            | 21-FEB-2001   |

|           |             |
|-----------|-------------|
| ACCESSION | AZ873604    |
| VERSION   | AZ873604.1  |
|           | GI:13081865 |

SOURCE ORGANISM

|                      |  |
|----------------------|--|
| REFERENCE<br>AUTHORS | 1 (bases 1 to 607)<br>Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,<br>Henderson, N., Macdonald, P., McDermott, T., Patterson, R. |
|----------------------|--|

and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10Kb

**JOURNAL**      **Unpublished (2000)**  
**COMMENT**      **Contact: Robert B. Weiss**

University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801/585 5506

Email: [ddunn@genetics.wash.edu](mailto:ddunn@genetics.wash.edu)  
 Insert Length: 1000 Std Error: 0.00  
 Plate: 0187 Row: J Column: 24  
 Seq primer: CGTCTAAMCGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 607.  
 Location(s): 1158

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source   | 1. .607             |

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db-xref="taxon:10090"
/clone="U06C2K0187J24"
/clone_lib="Mouse 10Kb plasmid U06C1M library"

```

*1*Ab<sub>2</sub> host E. coli strain XL10-Gold, T1-resistant, F<sup>-</sup>/note="vector: pMD42ny). Purified genomic DNA from *M. musculus* C57Bl/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The ligated DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (911473211419b) (AT29072-1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells



BASE COUNT 181 a 154 c 128 g 144 t  
 and selected for ampicillin resistance."

Query Match 60.0%; Score 15; DB 251; Length 607;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 gtcaggtgcaggtca 23  
 |||

Db 487 gtcaggtgcaggtca 473

RESULT 42  
 A2418216 624 bp DNA GSS 03-OCT-2000  
 LOCUS 1M0194P10F Mouse 10kb plasmid UUCG1M library Mus musculus genomic  
 DEFINITION clone UUCG1M0194P10 F. DNA sequence.  
 A2418216  
 VERSION A2418216.1 GI:105422229  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 624)  
 Authors: Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.,  
 and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Km. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0194 row: P column: 10  
 Seq primer: CGTGTAAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 624.

#### FEATURES

Location/Qualifiers  
 1..624  
 /organism="Mus musculus"  
 /strain="C57BL/6j"  
 /db\_xref="taxon:10090"  
 /clone="UUCG1M0194P10"  
 /clone\_lib="Mouse 10kb plasmid UUCG1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PMD42 (g14732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into

BASE COUNT 158 a 145 c 153 g 168 t  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

Query Match 60.0%; Score 15; DB 243; Length 624;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 gtcaggtgcaggtca 23  
 |||

Db 111 gtcaggtgcaggtca 97

RESULT 43  
 AW342762 639 bp mRNA EST 31-JAN-2000  
 LOCUS f186h10.x1 Sugano kawakami zebrafish DNA Danio rerio cDNA clone  
 DEFINITION 2644579.3 similar to TR:Q15312 Q15312 R KAPPA B.; mRNA sequence.  
 AW342762  
 VERSION AW342762.1 GI:6839128  
 KEYWORDS EST.  
 SOURCE zebrafish.  
 ORGANISM Danio rerio

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Rasbora; Danio.  
 1 (bases 1 to 639)  
 Authors: Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S.,  
 Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T.,  
 Martin, J., Page, D., Steptoe, M., Underwood, K., Theisling, B., Ritter,  
 E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE WashU Zebrafish EST Project 1999  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: S.L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu

Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA  
 Sequencing by: Washington University Genome Sequencing Center  
 Seq primer: T7 RT from Amersham  
 High quality sequence stop: 346.  
 Location/Qualifiers  
 1..639

#### FEATURES

Location/Qualifiers  
 1..639  
 /organism="Danio rerio"  
 /strain="AB"  
 /db\_xref="taxon:7955"  
 /clone="2644579"  
 /clone\_lib="Sugano Kawakami zebrafish DNA"  
 /sex="mixed (one male and one female, including  
 unfertilized eggs)"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (phage resistant)"  
 /note="Vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG);  
 Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed  
 with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT];  
 double-stranded cDNA was ligated to a DraIII adaptor  
 [TGTGGCCCTTACGCG], digested and cloned into distinct DraIII  
 sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site  
 CACCATGTG). XhoI should be used to isolate the cDNA  
 insert. Size selection was performed to exclude fragments  
 <1.5kb. Library constructed by Dr. Sumio Sugano  
 (University of Tokyo Institute of Medical Science) and  
 kindly donated by Dr. Koichi Kawakami. Custom primers for  
 sequencing: 5' end primer CTTCGTCTTAAAGCTGCG and 3' end  
 primer CGACCTGCGAGCTGACGACA."

BASE COUNT 127 a 146 c 192 g 174 t

ORIGIN

Query Match 60.0%; Score 15; DB 114; Length 639;

Best Local Similarity 100.0%; Pred. No. 1.7e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caggtcaggtcagc 25  
|||||

Db 438 CAGGTGAGGTGAGC 424

## RESULT 44

C98409

LOCUS C98409 658 bp mRNA EST 19-OCT-1998  
DEFINITION C98409 Rice panicle at flowering stage Oryza sativa cDNA clone

ACCESSION E0106\_62, mRNA sequence.

VERSION C98409.1 GI:3761161

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE 1 (bases 1 to 658)  
Sasaki, T. and Yamamoto, K.  
Rice cDNA from panicle at flowering stage

JOURNAL Unpublished (1996)

CONTACT: Takuji Sasaki  
National Institute of Agrobiological Resources

Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba

Ibaraki,  
Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468  
Email: tsasaki@affrc.go.jp

PROJECT = 'RGP'

## FEATURES

Source

Location/Qualifiers

1..658  
/organism="Oryza sativa"

/strain="Nipponbare"

/db\_xref="taxon:4530"

/clone="E0106\_62"

/clone\_1lb="Rice panicle at flowering stage"

/dev\_stage="flowering stage"

/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 161 a 159 c 155 g 179 t 4 others

ORIGIN

Query Match 60.0%; Score 15; DB 156; Length 658;

Best Local Similarity 100.0%; Pred. No. 1.7e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcctcgtcaggtgca 18  
|||||

Db 177 TCCTCGTCAGGTGCA 191

## RESULT 45

AUI65518

LOCUS AUI65518 668 bp mRNA EST 06-DEC-2000  
DEFINITION AUI65518 Rice panicle at flowering stage Oryza sativa cDNA clone

ACCESSION E2045, mRNA sequence.

VERSION AUI65518.1 GI:11564882

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehharitoidae; Oryzae; Oryza.

REFERENCE 1 (bases 1 to 668)

AUTHORS Sasaki, T. and Yamamoto, K.  
TITLE Rice cDNA from panicle at flowering stage (2000)  
JOURNAL Unpublished (2000)  
CONTACT: Takuji Sasaki  
National Institute of Agrobiological Resources

Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba

Ibaraki,  
Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasaki@affrc.go.jp

PROJECT = 'RGP'

E2045\_42.

## FEATURES

Source

Location/Qualifiers

1..668  
/organism="Oryza sativa"

/strain="Nipponbare"

/db\_xref="taxon:4530"

/clone="E2045"

/clone\_1lb="Rice panicle at flowering stage"

/dev\_stage="flowering stage"

/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 160 a 162 c 162 g 179 t 5 others

ORIGIN

Query Match 60.0%; Score 15; DB 108; Length 668;

Best Local Similarity 100.0%; Pred. No. 1.7e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcctcgtcaggtgca 18  
|||||

Db 198 TCCTCGTCAGGTGCA 212

Search completed: October 9, 2001, 15:15:39  
Job time: 13661 sec